# *Regulation of transcription in hyperthermophilic archaea*


Arjen Brinkman

# *Stellingen*

1. Het testen van alleen leucine als ligand voor een willekeurig Lrp eiwit is een waardeloos experiment.
   *Dit proefschrift.*
   *Enoru-Eta, J., D. Gigot, T.L. Thia-Toong, N. Glansdorff, and D. Charlier. 2000. Purification and characterization of Sa-Lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon Sulfolobus acidocaldarius homologous to the bacterial global transcriptional regulator Lrp. J Bacteriol 182: 3661-3672.*

2. Natuurlijk voorkomende bindingssequenties voor Lrp eiwitten zijn moeilijk aan te wijzen.
   *Dit proefschrift.*

3. Nieuwe technieken moeten niet als doel maar als middel gebruikt worden in wetenschappelijk onderzoek.

4. Voor acceptatie van een manuscript in een wetenschappelijk tijdschrift telt geluk minstens zo zwaar als de kwaliteit van het beschreven onderzoek.

5. Het gebruik van DNA-microarrays voor transcriptie analyse van hele genomen levert meestal meer vragen dan antwoorden op.

6. In de snelheid waarmee onderzoek gedaan kan worden speelt het behoorlijk functioneren van laboratorium apparatuur soms een grotere rol dan het tempo van de onderzoeker zelf.

7. Hoewel er in de politiek veel aandacht wordt besteed aan nationale veiligheid is de kans nog altijd groter een auto ongeluk te krijgen dan om te komen bij een terroristische aanval.

8. Misschien is niets geheel waar en zelfs dat niet.
   *Multatuli, Idee 1.*

Stellingen behorende bij het proefschrift getiteld:

***Regulation of transcription in hyperthermophilic archaea***

Arjen Brinkman, Wageningen, 8 augustus 2002

# *Regulation of transcription in hyperthermophilic archaea*

Arie Berdinus Brinkman

*Promotor:*

**prof. dr. W.M. de Vos**
Hoogleraar Microbiologie
Wageningen Univeriteit

*Co-promotor:*

**dr. J. van der Oost**
Universitair hoofddocent bij de leerstoelgroep Microbiologie
Wageningen Universiteit

*Leden van de promotie commissie:*

**dr. S.D. Bell**
The Hutchinson/MRC research center
Cambridge, Groot-Brittannië

**prof. dr. O.P. Kuipers**
Rijksuniversiteit Groningen

**prof. dr. H.G. Stunnenberg**
Katholieke Universiteit Nijmegen

**prof. dr. S.C. de Vries**
Wageningen Universiteit

# *Regulation of transcription in hyperthermophilic archaea*

Arie Berdinus Brinkman

**Proefschrift**

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
prof. dr. ir. L. Speelman,
in het openbaar te verdedigen
op vrijdag 6 september 2002
des namiddags te vier uur in de Aula.

**A.B. Brinkman** – Regulation of transcription in hyperthermophilic archaea– 2002
Dutch: 'Regulatie van transcriptie in hyperthermofiele archaea'

PhD Thesis Wageningen University, Wageningen, The Netherlands– With summary in Dutch

ISBN: 90-5808-673-9

*voor Mary*

The research described in this thesis was financially supported by the Council for Chemical Sciences (CW) of the Netherlands Organization for Scientific research (NWO) grant 700 35-101.

Cover design: Thilde Veltman

## *Dankwoord*

Mijn boekje is dan eindelijk af en de laatste regels zijn voor de mensen die mij op wat voor manier dan ook geholpen hebben. Veel mensen hebben hun steentje bijgedragen aan mijn project, en die wil ik daarvoor graag bedanken.

Hoewel ik de werkgroep Bacteriële Genetica binnenkwam als "import", heb ik mij de afgelopen jaren erg thuis gevoeld en heb ik met plezier bij de vakgroep Microbiologie gewerkt. John, jouw enthousiasme heeft mij enorm geholpen. Met jou als drijvende kracht achter het regulatieproject hebben we tijdens brainstorm sessies veel ideeën gegenereerd waarvan een aantal erg succesvol bleken. Daarnaast heb je me met datzelfde enthousiasme bij tijd en wijle net even het duwtje weten te geven om verder te gaan dan ik eigenlijk wilde. De mix van positivisme en een leuke samenwerking waren daarbij van grote waarde. Bedankt daarvoor! Willem, ook jouw enthousiasme over mijn onderzoek heb ik als erg motiverend ervaren. Tussen een druk programma door vond je toch altijd de tijd om het weer eens over wetenschap te hebben, mij de nodige suggesties te doen als het niet zo lekker liep en kritisch te zijn wanneer dat nodig was. Bedankt.

Alle mensen van Bacgen wil ik bedanken voor de leuke tijd: John, Ans (voor het sociale reilen en zeilen bij Bacgen), Servé, Ineke, Hauke, Judith, Joyce (voor de uurtjes schaatsplezier), Leon (voor de broodnodige klaagzang, goede humor, luchtgitaren en natuurlijk voor het paranimf zijn), Don (for American habits and good times at the Gordon conferences), Thijs K. (voor schaatsplezier en muzikale belevenissen), Corné (voor je hulp als paranimf, schaatsen, discussies over onderzoek en de fijne "ambiance" op de kamer). Gaël, Valerie, Ana, Ronnie, Krisztina, Pino, Nina en Alexandra, The Friends Of *Sulfolobus*: Thijs E., Stan, and Ken (for good discussions about our friend). De dagelijkse lunches bij Bacgen zal ik niet snel vergeten. De discussies waren meestal bijzonder, leuk, ontspannen en daarom een belangrijk ingrediënt van mijn promotieonderzoek. Ria, Nees, Wim en Francis, jullie wil ik bedanken voor alles wat er naast onderzoek komt kijken bij promoveren. Ook de studenten die hebben meegewerkt aan mijn onderzoek, Edwin, Jan, Guido, Robert Jan en Noor, jullie wil ik bedanken voor jullie inzet en bijdrage. Steve, your help with the project has been of immense value. Many thanks for answering countless questions, advice, inspiring discussions, and humor during two Gordon conferences, but in particular during my two visits to Cambridge. Although LysM did not behave as straightforward as we hoped for, the resulting paper is a proof of this very good collaboration. Isabell, thanks for your collaboration on the LrpA paper.

Iedereen die op wat voor manier dan ook heeft bijgedragen aan dit boekje wil ik bedanken voor de nodige afleiding, getoonde interesse of het aanhoren van frustraties. Haico, de vele squashavonden zijn hier zeker een voorbeeld van, al kwam ik door verloren wedstrijden niet echt van die frustraties af… Albert, je advies over mijn keuze wel of niet te gaan promoveren heb ik in de wind geslagen geloof ik. Sorry, maar toch bedankt voor tips en gezellige avonden. Heren van de band: Anton, Wim en Jan, onbewust hebben jullie voor veel ontspanning gezorgd. Vrienden uit Amerongen e.o., bedankt voor de gezellige en ontspannen dagen op Terschelling, Limburg, Drenthe en Oostenrijk. Mijn "micro" benadering kwam toch goed van pas! Thilde, bedankt voor je hulp tijdens het leggen van het ei dat omslag heet. Ben, Co, Conny, Frederieke, Henk, Miriam, Thilde en Geert, oom Jan, tante Betty, bedankt voor het luisterende oor en de gezelligheid. Papa en mama, jullie wil ik bedanken voor de kans die jullie me altijd hebben gegeven om te doen wat ik wilde, of het nu Conservatorium, HLO of de VS was. Zonder de mogelijkheid deze vreemde kronkelweg te volgen was dit boekje er zeker niet geweest.

Hidde, jouw komst en die van dit proefschrift vielen wat vreemd samen. Je zorgde daarbij voor de nodige afleiding, wat je naar twee kanten mag uitleggen. Ik geniet met volle teugen van jou, en ik hoop dat we nog lange, lange tijd maatjes kunnen zijn!

Als laatste wil ik jou, Mary, bedanken voor je aandacht, liefde, vriendschap en je opofferingen. Jouw inzet en aandacht waren onmisbaar. Het eind van deze periode is voor ons tevens het begin van een hele mooie nieuwe, waarin we zeker voor nieuwe uitdagingen komen te staan. Ik hoop dat we deze samen kunnen nemen, en dat we nog lang van elkaar kunnen genieten!

## *Table of contents*

# *Chapter 1*

General introduction

## THE ARCHAEAL DOMAIN

In the late 1970s phylogenetic analyses of small-subunit rRNA sequences showed that there is a principal division within the prokaryotic domain (138), separating the then-called eubacteria from the archaeabacteria. Woese and coworkers therefore proposed a three-domain division among all living organisms, consisting of bacteria, archaea and eukarya (139) (Fig. 1). The domain of the archaea can be further divided into two kingdoms: the euryarchaeota and the crenarchaeota. Since the recognition of archaea as a distinct phylogenetic domain of life, many studies have focused on the physiology and molecular biology of archaea. First, because of the fundamental interest in unique archaeal biochemistry and cellular processes (e.g. ether lipid-containing membranes, methanogenesis, and transcription). Second, because of the potential application of relative robust archaeal enzymes. Many archaea are so-called extremophiles because of their ability to grow under various extreme conditions regarding salt concentration (halophiles), pressure (barophiles), pH (acidophiles or alkaliphiles), or temperature (psychrophiles or thermophiles). Enzymes from extremophilic archaea are usually intrinsically stable under the harsh conditions of their habitat (124).

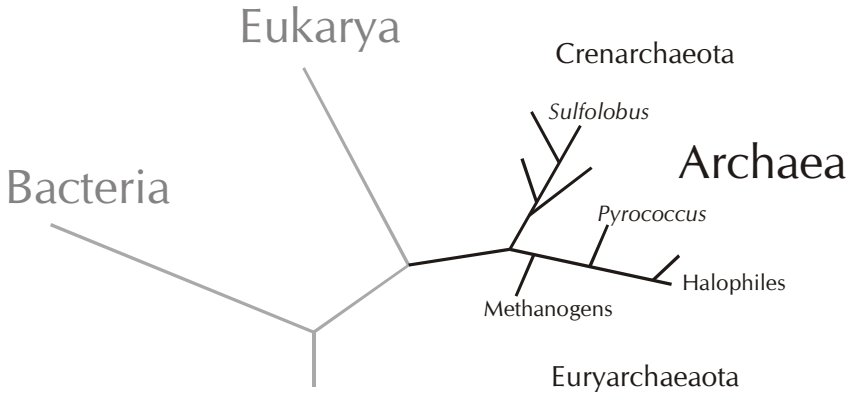



**Figure 1.** Universal tree of life, based on phylogenetic analysis of 16S/18S rRNA sequences (139). The three domains bacteria, eukarya, and archaea are depicted. The archaeal domain can be further divided into two kingdoms: the crenarchaeota and the euryarchaeota.

Organisms that have an optimal growth temperature of 80°C or higher are called hyperthermophiles. The hyperthermophilic archaea *Pyrococcus furiosus* and *Sulfolobus solfataricus* are the key organisms in this thesis. *P.furiosus* is a euryarchaeotic marine organism that was isolated from geothermally heated sediments at the island Vulcano, Italy (47). The organism grows strictly anaerobically at an optimal temperature of 100°C, with a doubling

time of about 30 min.. *S.solfataricus* is a crenarchaeote, isolated from a solfataric field near Naples, Italy. It is an aerobic thermoacidophile, growing optimally at 80°C and pH 2-4, with a doubling time of several hours (146). Both organisms are versatile organisms, able to grow heterotrophically on a variety of sugars and proteinaceous substrates. The genomes of *P.furiosus* and *S.solfataricus* have been completely sequenced. They are 1.9 and 3.0 Mbp in size, and display a GC-content of 41% and 38%, respectively (Utah Genome Center, Dept. of Human Genetics, University of Utah, http://www.genome.utah.edu) (119).

## THE ARCHAEAL TRANSCRIPTION MACHINERY

Although archaea resemble bacteria with respect to their prokaryotic cellular organization, the biochemistry of some of their fundamental cellular processes like transcription, translation, and replication is remarkably different (15, 91). The archaeal transcription machinery does not resemble that of bacteria, but is rather eukaryal-like. The first step in transcription initiation is the assembly of the preinitiation complex (PIC) at the promoter. A schematic comparison of the PICs from the three domains of life is shown in Fig. 2. Although eukarya have three distinct RNA polymerase (RNAP) systems, only the type II system is shown, since this is most comparable to the archaeal system (see below). Formation of the eukaryal type II PIC is accomplished by the interaction of multiple general transcription factors (114), starting with the binding of the multi-subunit complex TFIID to a promoter element called the TATA box. TFIID consists of TATA-binding protein (TBP) that is responsible for TATA-box binding, and TBP associated factors (TAFs). The binding of TFIID can be stabilized by TFIIA. TFIIB subsequently binds the TFIID-DNA complex and recruits the RNA polymerase-TFIIF complex. The binding of TFIIE and TFIIH further completes the PIC (114).

*In vitro* transcription systems have been developed for several archaea (39, 48, 61, 66, 69, 106), allowing investigation of archaeal requirements for specific transcription. Comparison of the eukaryal with the archaeal PIC shows remarkable similarities. The archaeal PIC includes TBP, TFB (orthologous to TFIIB), and a eukaryal-like RNAP (see below). Hence, the archaeal PIC appears to be a simplified version of the eukaryal RNAP II system. Interestingly, a subset of eukaryal promoters can accurately be transcribed using a minimal transcription complex, consisting of only TBP, TFIIB, the RAP30 subunit of TFIIF, and RNAP (129). This shows that a eukaryal transcription system with comparable simplicity to the archaeal system also has basic transcriptional activity.

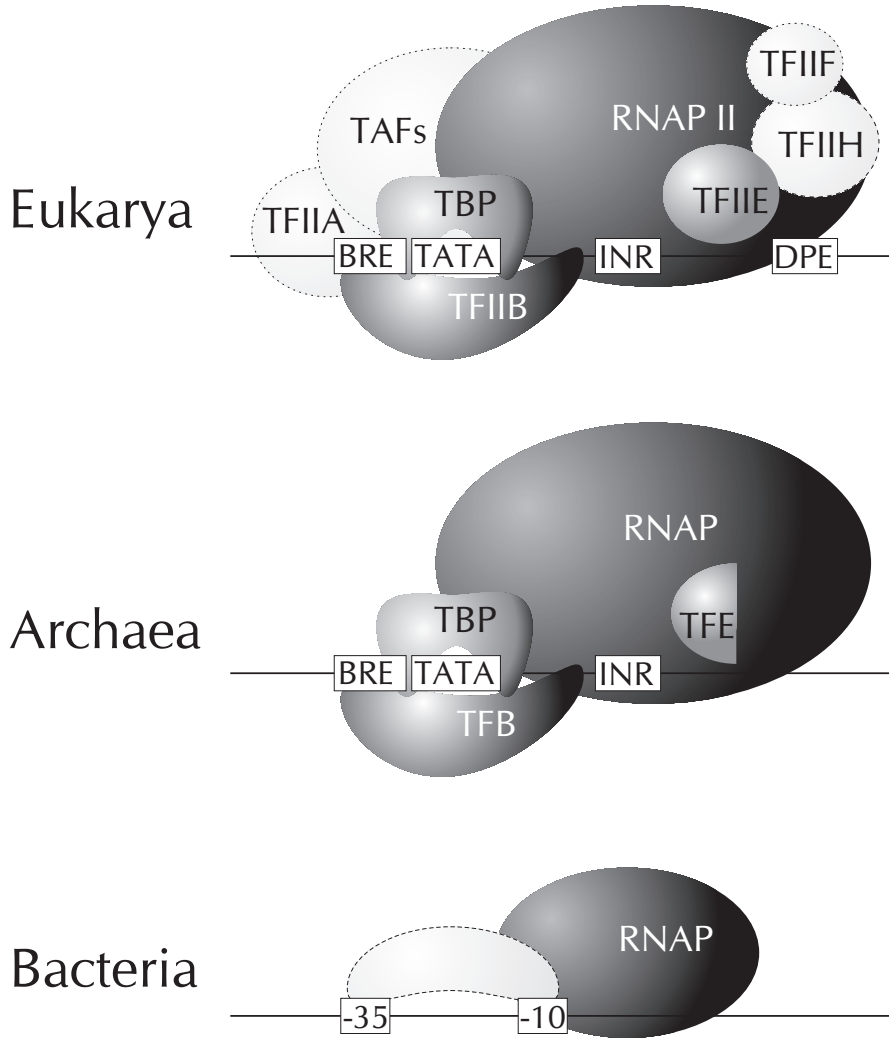



**Figure 2.** Schematic representation of pre-initiation complexes and promoters in the three domains of life. Objects with dashed lines indicate proteins absent in archaea. See text for further explanation.

The bacterial PIC is clearly different from that of archaea and eukarya. First, the bacterial RNAP is smaller, the core enzyme consisting of five subunits in a $\alpha_2\beta\beta'\omega$ configuration (141). Second, the bacterial RNAP holoenzyme, consisting of the core enzyme and a σ-factor, has the intrinsic ability to specifically recognize core promoter elements of bacterial promoters, usually located at -35 and -10 bp upstream of the transcription start site. The core RNAP is able to combine with one of the several different σ-factors, each of which enables the RNAP holoenzyme to direct transcription from a specific subset of promoters. For example, *E.coli* utilizes at least five different σ-factors, whereas no less than 18 σ-factors are encoded by the *B.subtilis* genome (56, 77). No

additional transcription factors are necessary for promoter recognition and initiation of transcription. Since σ-factors are absent in both eukarya and archaea, they appear to be typical for the bacterial transcription system.

## RNA polymerase

*S.acidocaldarius* RNAP (saRNAP) was among the first archaeal RNAPs purified, and the subunit complexity and sequences revealed that it is more related to eukaryal than to bacterial polymerases (82, 145). Archaeal RNAPs have at least ten subunits (Table I). Their large subunits share homology with both bacterial and eukaryal RNAP large subunits, but they have the highest homology with those of eukarya (82). More specifically, they are most closely related with the RNAP II subunits, as determined by phylogenetic analysis (15). Importantly, however, the archaeal RNAP subunit A" lacks the characteristic heptad repeats found in the C-terminal part of the homologous eukaryal largest subunit RPB1, where these are the targets of extensive phosphorylation after transcription initiation (36). The SaRNAP subunits E, H, I, and N have homology only with eukaryal RNAP subunits. Among the different archaea there are variations in RNAP subunit organization. Subunit B is sometimes split into B' and B" (methanogens and halophiles), and subunit E is split into E' and E" (methanogens and *S.solfataricus*) (14, 144). Furthermore, a comparison of SaRNAP with *Methanobacterium thermoautotrophicum* RNAP (MtRNAP) showed that the SaRNAP subunits G, M, and possibly N are not present in MtRNAP, and in contrast, MtRNAP subunit P is absent in SaRNAP (39).

Interaction studies have shown that archaeal and eukaryal RNAP subunits are not only comparable at the amino acid sequence level, but also at the functional level. The archaeal D and L subunits interact, as do their eukaryal homologues RPC5 and RPC9 (18, 81). In addition, archaeal D is capable of interacting with eukaryal RPC9 (41). All these subunits contain motifs known as "alpha motifs" from α subunits of bacterial RNAP that are involved in the homo-dimerization of the two α subunits. Additional interactions have been shown for archaeal F-E and P-D, and their eukaryal homologues RPB4-RPB7 and RPB12-RPB3, respectively (136). Again, an archaeal-eukaryal F-RPB7 interaction can be detected as well. The recently solved crystal structures of *S.cerevisiae* RNAP and *T.aquaticus* RNAP allow for an even more detailed view of the location and interactions of the different subunits within the RNAP (35, 141). In conclusion, interactions that take place between eukaryal RNAP subunits also occur between the homologous archaeal subunits, and some of the involved subunits are functionally interchangeable (D-RPC5 and F-RPB7), at least regarding the interaction properties. It demonstrates a functional relationship

among archaeal and eukaryal RNAP subunits, and indicates conservation across the evolutionary domain boundary.

| RNAP subunit of *M.thermoauto-trophicum* | RNAP subunit of *S.acidocaldarius* | Homologous RNAP subunits | | | |
|---|---|---|---|---|---|
| | | Bacterial (*E.coli*) | Eukaryal (*S.cerevisiae*) | | |
| | | | I | II | III |
| A′ A′′ | A′ A′′ | β′ | RPA1 * | RPB1 * | RPC1 * |
| B′ B′′ ¶ | B | β | RPA2 | RPB2 | RPC2 |
| D | D | α | RPC5 | RPB3 | RPC5 |
| E' | E | - | - | RPB7 | RPC25 |
| F | I | - | - | RPB4 | - |
| - | G | - | - | - | - |
| H | H | - | RPB5 § | RPB5 § | RPB5 § |
| K | K | ω ‡ | RPB6 # | RPB6 # | RPB6 # |
| L | L | α | RPC9 | RPB11 | RPC9 |
| - | M | - | - | - | - |
| - | N | - | RPB10 | RPB10 | RPB10 |
| P | - | - | RPB12 | RPB12 | RPB12 |

**Table I**. Archaeal RNAP subunits and their corresponding homologues in bacterial and eukaryal RNAPs. Adapted from Langer *et al*. (82), Darcy *et al*. (39), and Bell & Jackson (11, 14). ¶*M.thermoautotrophicum* B′ and B′′ are homologous to the N-terminal and C-terminal 1/2 of *S.acidocaldarius* B, respectively**;** *A′ and A′′ are homologous to the N-terminal 2/3 and the C-terminal 1/3, respectively; §C-terminal 1/3; #C-terminal 1/2; ‡ω is present in the core and holoenzymes of bacterial RNAPs, and is encoded by all available bacteria genomes. It is, however, not essential for transcription *in vivo* and *in vitro* (24, 50, 64, 141).

## TBP

Archaeal TATA-binding proteins (aTBPs) are homologous to eukaryal TBPs (eTBPs). Although aTBPs do not contain the N-terminal region of eTBPs that is highly variable in sequence and length, their core domains are about 40% identical (15) (Fig. 3A). Some aTBPs have a C-terminal stretch of six to ten highly acidic amino acid residues, but the function of these residues is unknown (14). The homologous core domains of both eTBPs and aTBPs are composed of imperfect direct repeats, however, the symmetry between the direct repeat elements in aTBPs is higher than in eTBPs (40% vs. 28-30% identity, respectively) (14). The symmetry of these domains is also evident in the solved three-dimensional structure from *Pyrococcys woesei* TBP (40). The structure of this aTBP is very similar to that of eTBPs (98) (Fig. 3B). Both aTBP and eTBPs have a characteristic "saddle shape". Their concave inner DNA-binding side consists of an antiparallel β-sheet, while a "stirrup loop" is present at each end of the long axis. eTBP is required for transcription by

all three eukaryal RNAP systems, and is part of one of the three multi-subunit complexes SL1, TFIID, or TFIIIB, specific for RNAP I, II, or III, respectively. Although aTBP exists as a homodimer in archaeal cell extracts like TFIID, it is not complexed with any other protein (106, 126). Moreover, archaeal genome sequences do not contain any homologues of TBP-associated factors (TAFs), the proteins associated with TBP in the eukaryal SL-1, TFIID, and TFIIIB complexes.

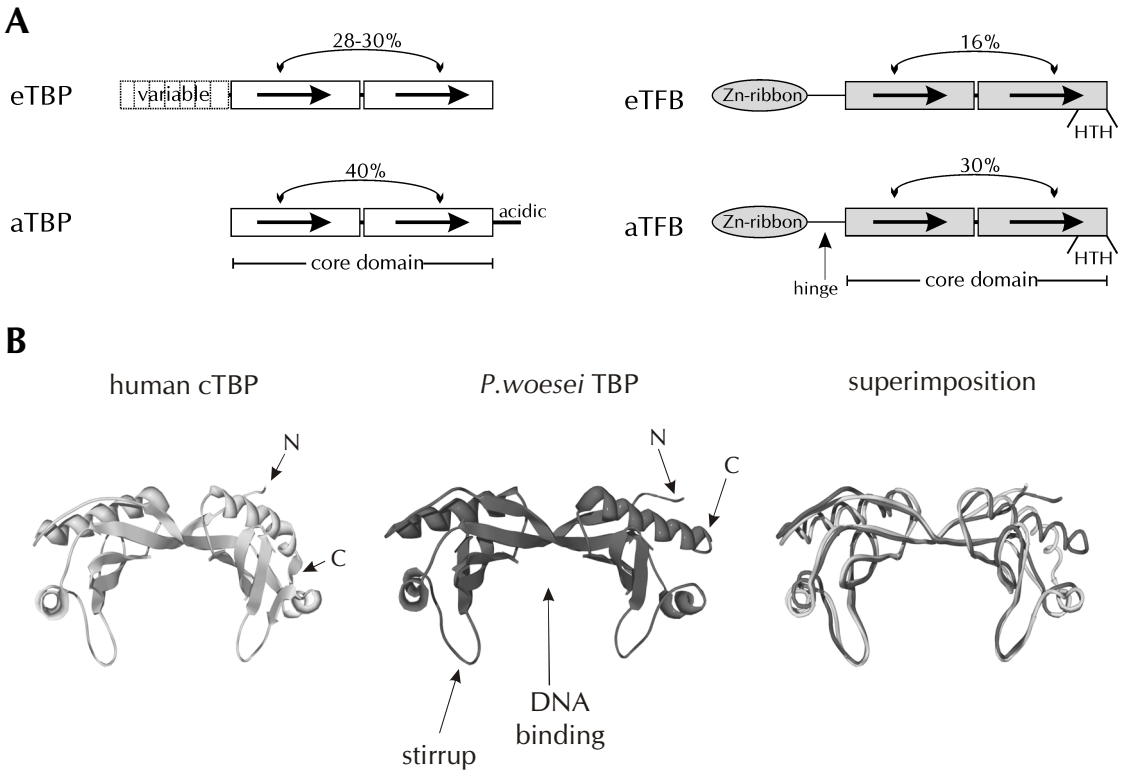



**Figure 3.** Structure of TBP and TFB. (**A**), Schematic representation of eukaryal TBP (eTBP), archaeal TBP (aTBP), eukaryal TFB (eTFB), and archaeal TFB (aTFB). eTBPs contain an N-terminal extension of variable length, whereas some aTBPs contain a short C-terminal acidic region. Horizontal arrows indicate imperfect direct repeats in amino acid sequence, and the percentage of identity is indicated. (**B**), Three-dimensional structure of the human core domain of TBP (98) and *P.woesei* TBP (40). N-termini and C-termini are indicated, as well as the antiparallel DNA-binding β-sheet and the characteristic "stirrup" loop. The structures of human cTBP and *P.woesei* TBP are superimposed to indicate their structural homology.

The essential role of aTBP in archaeal transcription was demonstrated by immunodepletion of aTBP from *Sulfolobus shibatae* cell extracts. No transcription activity was present in these extracts, but this could be restored by addition of recombinant *S.shibatae* aTBP (105). In addition, aTBP can be replaced by eTBP from either yeast or human, which both facilitate specific transcription in a *Methanococcus thermolithotrophicus in vitro* transcription system (137).

Thus, eukaryal and archaeal TBPs are structurally and functionally comparable, and nucleate assembly of the PIC.

## TFB

The second archaeal transcription factor is transcription factor B (TFB), which is homologous to the eukaryal RNAP II-specific transcription factor IIB (TFIIB). Since archaea have only one RNAP system, their factor B is simply referred to as TFB. Eukaryal TFIIBs and archaeal TFBs are about 30% identical (14), and share structural homology (74). Both proteins consist of two domains: an N-terminal Zn-ribbon domain and a larger C-terminal core domain consisting of an imperfect direct repeat (Fig. 3A). Like TBP, TFB is an essential archaeal transcription factor (66, 106), which has been shown to be involved in a series of events leading to transcription initiation, as is detailed below.

First, by binding of TFB to the TBP-DNA complex, a more stable ternary TFB-TBP-DNA complex is formed, which in turn is able to recruit RNAP. The X-ray crystal structure of the *P.woesei* ternary complex containing the core domain of TFB (cTFB) showed that TFB binds the DNA helix opposite of TBP, interacting with the C-terminal stirrup loop of TBP (74, 86) (Fig. 4). The structure of this archaeal ternary complex is highly comparable to the eukaryal ternary DNA-TBP-cTFIIB complex, again indicating that the transcription systems in the two domains are very well conserved.

Second, TFB determines transcriptional polarity. It was shown that a sequence immediately upstream of the TATA box, called the TFB-responsive element (BRE), is important for TFB-mediated promoter strength *in vitro* (66, 107) (Fig. 2). Placement of the BRE downstream of the TATA box instead of upstream resulted in a reversed direction of transcription (17). Inversion of the TATA box while leaving the BRE unaffected had no effect on transcription direction (17), which is in agreement with the observation that TBP is a highly symmetric protein (see above), possibly capable of binding the TATA box in both orientations, like eTBP (34). Subsequently, it was shown that there is indeed a direct sequence-specific interaction between TFB and the BRE (17, 107). The X-ray crystal structure of the ternary complex confirmed these results, and showed that a C-terminal helix-turn-helix (HTH) motif of TFB makes base-specific contacts with the BRE (Fig. 4) (86).

Third, TFB recruits RNAP. Although the N-terminal domain of TFB is dispensable for formation and stabilization of the ternary complex, it is crucial for RNAP recruitment (13). The structure of this domain resembles that of a Zn-ribbon (Fig. 7B) (142), a fold usually involved in protein-DNA or protein-protein interactions. This domain was found

to interact primarily with the RpoK subunit of saRNAP, a subunit that is conserved in all three eukaryal RNAPs (RPB6), as well as in bacterial RNAP (ω) (92). The corresponding N-terminal domain of eukaryal TFIIB has also been shown to involved in RNAP recruitment (23), however, eukaryal RNAP recruitment appears to be mediated by TFIIF, which has not been identified in any of the archaeal genomes.

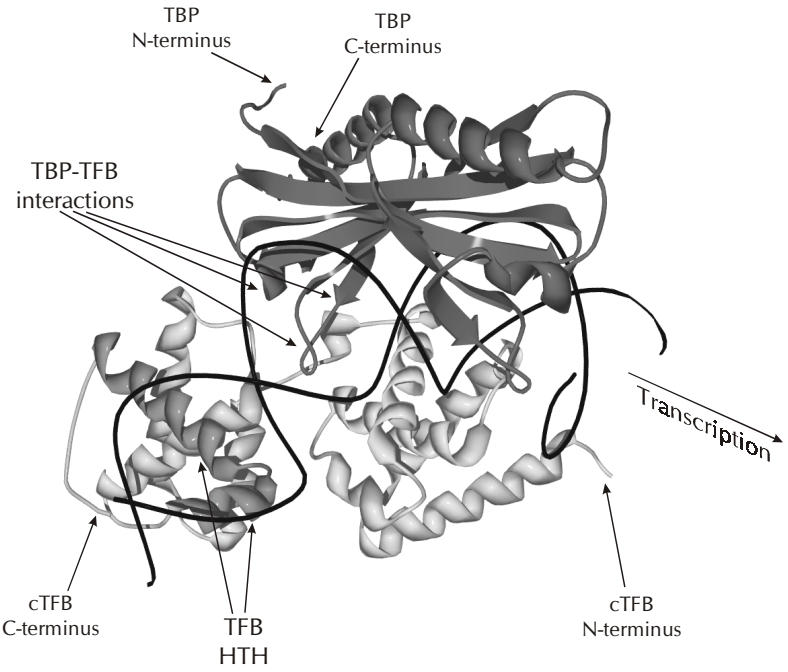



**Figure 4.** Three-dimensional structure of the *P.woesei* ternary complex consisting of TATA-BRE-DNA (black), TBP (gray) and the core domain of TFB (white) (86). N-termini and C-termini of the proteins are indicated, as well as the direction of transcription, the C-terminal HTH DNA-binding domain of TFB (darker helices), and the TBP-TFB interface (arrows).

Apart from multiple roles in early stages of transcription initiation, TFB also plays a role in promoter clearance, a late stage in transcription initiation (13). Mutation of a conserved glutamate residue among TF(II)Bs (E46), located in a presumable flexible hinge region between the Zn-ribbon and the core domain, abrogates clearance at some promoters. Although the effect appears to be dictated by the DNA sequence surrounding the transcription start site, the exact mechanism by which TFB affects promoter clearance is not known yet.

It is interesting to note that some archaea encode multiple copies of TBP or TFB. For instance, *Halobacterium sp.* NRC-1 contains at least five genes encoding TBP paralogues and seven genes encoding TFB paralogues, either on its plasmid or its chromosomal DNA (97). It is tempting to speculate that these different TBPs or TFBs facilitate transcription from

specific subsets of promoters, in analogy to the function of bacterial sigma factors (see also below).

## Other transcription factors

The *in vitro* transcription of many tested archaeal promoters has been demonstrated to be independent of transcription factors other than TBP and TFB, and RNAP. However, archaeal genomes encode two additional eukaryal-like transcription factors, TFS and TFE (homologous to TFIIS and TFIIEα, respectively), both of which have been studied *in vitro*. As described in Chapter 2, TFE plays a stimulatory role on some promoters and under certain conditions. TFS has homology to both the eukaryal transcription elongation factor TFIIS and small subunits of RNAP I, II, and III, but was shown to be not complexed with archaeal RNAP. TFS induces a 3' to 5' transcript cleavage activity in RNAP, resulting in the release of stalled elongation complexes, and its function is therefore similar to that of eukaryal TFIIS (62).

Homologues of other eukaryal type II general transcription factors, like TFIIA, TFIIF, and TFIIH are not encoded by archaeal genomes. The absence of an archaeal TFIIH homologue appears to be logical: eukaryal TFIIH is involved in phosphorylation of C-terminal heptad repeats of the RNAP largest subunit and ATP-dependent promoter melting (114). The archaeal RNAP large subunit does not contain these repeats, and does not require ATP hydrolysis for transcription initiation (16, 63).

## Archaeal promoters and terminators

In analogy with the presence of a eukaryal-like RNAP and transcription factors in archaea, promoter elements were also found to be eukaryal-like, resembling RNAP II promoters and a subclass of RNAP III promoters (93, 110). Extensive mutational analyses of archaeal promoters *in vitro* and *in vivo* allowed the definition of an archaeal consensus promoter (38, 100, 110), which consists of three major elements: the TATA box, the BRE, and the initiator element (INR, see Fig. 5). An *in silico* analysis of mapped archaeal promoters confirmed the experimental mutational data and showed a non-random base distribution at positions corresponding to TATA, BRE, and INR (122). In addition, this analysis showed that consensus sequences vary between halophilic archaea, methanogenic archaea, and crenarchaeota. A comparison of archaeal promoter sequence elements is shown in Fig. 5. The TATA box is the target site for TBP, and its name refers to an often-encountered alternating sequence consisting of As and Ts, where Cs and Gs are avoided, except at the first position (122).

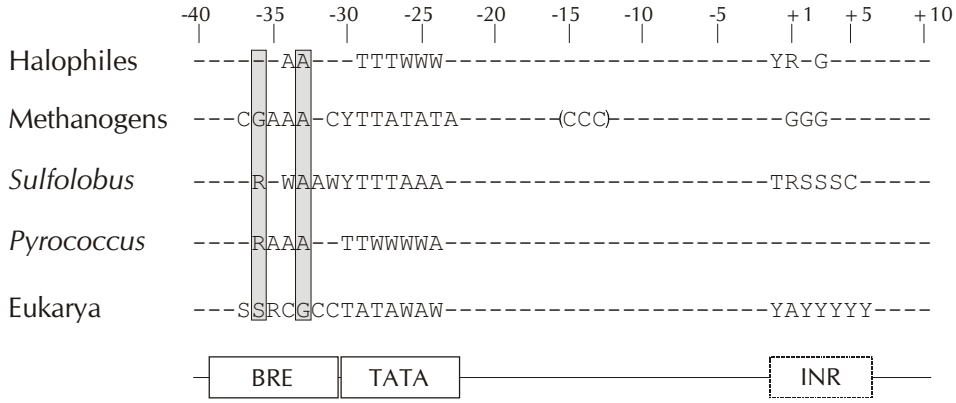



**Figure 5.** Consensus sequences of archaeal promoters. BRE, TFB-responsive element; INR, initiator element. W = A/T; Y = C/T; R = A/G; S = G/C; N = A/C/G/T. Numbers indicate the position relative to the transcription start (+1). For comparison, the consensus sequences of eukaryal core promoter elements is given (19, 80). Key positions of the BRE involved in base-specific TFB-DNA interactions are boxed in gray. The size of the TATA-box and the BRE is eight and nine bp, respectively, as determined from structural studies, although this is not always reflected by sequence preference. The exact size of the INR is unknown, as indicated by the dashed box. Data was obtained from Soppa (122), Bell *et al*. (17), and Verhees (131).

There is some flexibility in the spacing between the TATA-box and the transcriptional start site, so that the center of the TATA box is located at 25-30 basepairs upstream of the transcriptional start site.

The BRE is recognized by a C-terminal HTH motif of TFB (Fig. 4). Although BRE sizes appear to differ among the archaeal subgroups (122), they generally consist of a six-bp purine-rich sequence (17), located directly upstream from the TATA box. Purines at positions –3 and –6 are key determinants for specific BRE recognition, since they are the targets for base-specific interaction with the HTH of TFB (86). The function of the TATA box and the BRE in both archaea and eukarya are comparable (80).

Although several studies have demonstrated the presence of the archaeal INR element (110, 122), its exact role has not been elucidated yet. In eukarya TFIID binds the INR (114, 120), however, this interaction is through one of the TAFs and not through TFB. In addition, the INR-TFIID interaction is enhanced by TFIIA. Since there are no obvious archaeal homologues of TAFs or TFIIA, the archaeal INR might function in a different way compared to that of eukarya. Because the E46K mutation in TFB sensitizes transcription initiation to sequence mutations in the INR (13) (see above), it is not unlikely that TFB is somehow related to INR function. It should be noted, however, that eukaryal RNAP has been shown to recognize transcription start sites (135).

In contrast to all of the above described promoter elements, the downstream promoter element (DPE) is only present in eukarya, and not in archaea. It is a 7-bp element that is

present in many TATA-less promoters, centered at position +30 (25). Like the INR element of eukarya, it is recognized by TFIID, not through TBP but through the TAFs, which are absent from archaea. In the current model, the eukaryal DPE functions as a downstream analogue of a TATA box, thus enabling TFIID binding at the promoter (25).

In comparison with the mechanism of transcription initiation of archaea, there is relatively little information about transcription termination. Experiments in which 3' termini of archaeal mRNA have been mapped suggest that transcription terminates at multiple positions within one sequence region or two closely spaced regions (37). Only in a few cases the termination site has been mapped to a single nucleotide or dinucleotide. In general, 3' termini map within poly-T stretches ranging from four to 30 nucleotides and an average length of about 15 nucleotides. Inverted repeats are only present in some experimentally determined 3' ends, but the predicted secondary structures produced by these sequences have little in common. In addition, a computer-based analysis of complete prokaryotic genomes showed that the hairpin-forming potential around the stop codons of archaeal genes is very low, which confirms the experimental data mentioned above (134). The absence of obvious hairpin-forming sequences suggests that it is unlikely that archaeal terminators resemble the Rho-independent terminators of bacteria. Although no obvious homologues of bacterial Rho are encoded by archaeal genomes, a variety of archaeal genomes encode homologues of the bacterial termination factors NusA and NusG. However, no experimental work has been reported on the characterization or function of these proteins in archaea.

## Topology of the DNA template

Whereas the DNA of bacteria, eukarya, and mesophilic archaea is negatively supercoiled, the DNA of hyperthermophilic archaea ranges from relaxed to positively supercoiled (26, 87). This positive supercoiling is thought to prevent denaturation of the DNA due to the high growth temperatures of hyperthermophilic archaea. Besides, these organisms were shown to modulate the topology of their DNA in response to thermal stress conditions, i.e. heat shock causes (transient) positive supercoiling, whereas cold shock causes negative supercoiling (88).

Since transcription requires local melting of DNA, it can be expected that the topological state of the DNA template affects the efficiency of transcription, and perhaps acts as a global regulatory mechanism. In fact, global changes in supercoiling can act locally, by regulating supercoiling-sensitive promoters (99, 133), a phenomenon that has also been demonstrated in archaea (6). The effect of DNA topology on transcription has been

investigated using *in vitro* transcription systems of *S.solfataricus* and *P.furiosus* (16, 65). *S.solfataricus* transcribes negatively to positively supercoiled DNA at physiological temperatures, but only negatively supercoiled to relaxed DNA at lower temperatures, and the block to transcription of positively supercoiled DNA at lower temperatures is at the level of RNAP recruitment and promoter opening (16). These results are in agreement with the observed induction of negative supercoiling upon cold shock in *Sulfolobales* (88), which possibly serves as a global response that ensures efficient transcription at sub-optimal temperatures (16). In marked contrast to *S.solfataricus*, *P.furiosus* transcribes only negatively supercoiled to relaxed DNA at physiological temperatures, and negatively to positively supercoiled DNA at lower temperatures. It is unclear whether this difference depends on the type of promoter or on a particular property of one of the two systems (65).

## Chromatin structure

The structure of the chromatin template plays a major role in eukaryal transcriptional regulation. Instead of regarding the DNA template as naked DNA at which the PIC freely assembles, the large genomes of eukarya are efficiently packaged, and therefore the template for transcription should rather be regarded as protected or buried by histones, together called chromatin. A variety of post-translational histone modifications and remodeling events commonly precedes eukaryal PIC formation, and these mechanisms are usually subject to complex regulatory processes (73). Although prokaryotes have much smaller genomes than eukarya, they also compact their DNA. For this purpose bacteria have histone-like proteins such as HU and H-NS (104). While these proteins are not as actively involved in regulation of gene expression as in eukarya, they do sometimes play a role (103).

Homologues of the eukaryal histones are also present in the euryarchaeota. Although the small archaeal genomes presumably do not require the rigorous packaging mechanism encountered in eukarya, archaeal histones form nucleosomes that are well comparable to the eukaryal core nucleosomes. The four eukaryal histones H2A, H2B, H3, and H4 have a core domain which forms the "histone fold" (4), and archaeal consensus histone sequences are homologous to this core domain (4, 115, 123). Archaeal histones lack the typical N-terminal and C-terminal extensions of eukaryal histones. These extensions are not strictly required for nucleosome assembly, but rather form the targets for interaction with regulatory proteins, acetylation, or are involved in higher-order chromatin assembly (73, 125). In general, dimerization of histones stabilizes the histone fold (70), but whereas dimer formation in eukarya is restricted to the formation of (H2A+H2B) and (H3+H4)

heterodimers, archaeal histones form homodimers as well as heterodimers (52). Because the number of histone-encoding genes varies from two to five per archaeal genome, it appears that formation of a variety of heterodimers is possible, potentially displaying various biological activities during different growth conditions (52). Two eukaryal (H3+H4) dimers assemble to form a $(H3+H4)_2$ tetramer which recognizes nucleosome positioning signals on the DNA. Subsequently an (H2A+H2B) dimer binds on each side of the $(H3+H4)_2$ tetramer to complete the nucleosome, which wraps 146 bp of DNA around this core (90). The archaeal nucleosome consists of a tetramer rather than an octamer (51, 101), and it resembles the structure that is formed by the $(H3+H4)_2$ tetramer core at the center of the eukaryal nucleosome (102) (Fig. 6). Although the eukaryal octameric nucleosome induces only negative supercoiling of its DNA, it has been shown that the $(H3+H4)_2$ core tetramer is able to wrap DNA into either negative or positive supercoils (58, 59). In analogy, the archaeal tetrameric nucleosome can induce both negative and positive supercoiling, depending on protein-DNA ratio or salt concentration (94, 95, 109). Since negative supercoiling is observed under more physiological high-salt conditions, is it likely that this reflects the *in vivo* situation (94).

The effect of histone binding to a DNA template transcribed *in vitro* was tested for the *P.furiosus gdh* promoter (121). It was shown that transcription was reduced by histone binding, in a manner that was dependent on the histone-to-DNA ratio and on the DNA template topology. Sensitivity of transcription to histones was highest with relaxed circular DNA, lower with negatively supercoiled DNA, and lowest on linear DNA. While *gdh* transcription *in vitro* is dependent on template topology, with negatively supercoiled DNA being a better template than relaxed DNA (65), it is likely that the repressive effect of histones is the result of a change in the DNA geometry (121). In addition, binding of histones to the DNA template presumably provides a physical barrier for transcription.

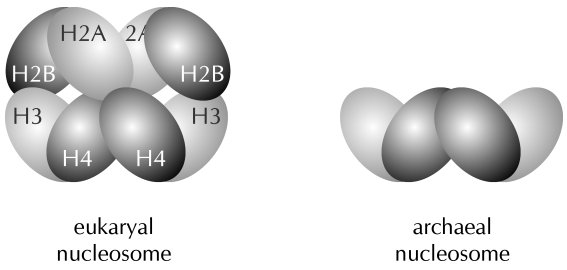



**Figure 6.** Schematic representation of the organization of histones within the eukaryal and archaeal nucleosome. The eukaryal nucleosome consists of an octamer containing two (H3+H4) dimers and two (H2A+H2B) dimers. The archaeal nucleosome resembles the core eukaryal $(H3+H4)_2$ tetramer, and homodimers or heterodimers are formed by any combination of the present archaeal histones.

Although histones appear to be restricted to the euryarchaeal kingdom, not all euryarchaeal species actually possess histones. For instance, *Thermoplasma* species instead

contain a protein called HTa, which is homologous to bacterial HU, and in *Methanosarcinaceae* MC1 is the most abundant DNA-associated protein (27-29).

Histones have not been detected in crenarchaeota. Instead, crenarchaeota have three classes of abundant DNA-binding proteins with molecular weights of 7, 8, and 10 kDa (31). Sac7d and Sso7d from *S.acidocaldarius* and *S.solfataricus*, respectively, are the major representatives of the 7-kDa class. Sso7d is a high-affinity DNA-binding protein which is *in vivo* methylated at five out of 14 lysine residues, the exact function of which is not yet known (31). Apart from protection of DNA against melting at high temperatures (57), binding of Sso7d leads to negative supercoiling of its bound DNA, and it is hypothesized that this Sso7d-induced unwinding of DNA plays a role during thermal stress. Presumably, Sso7d is therefore a functional analogue of the protein HU from *E.coli* (89).

## ARCHAEAL TRANSCRIPTIONAL REGULATORS

Whereas the basal transcriptional machinery of archaea is eukaryal-like, this is not the case for the regulatory proteins encoded by the archaeal genomes. Instead, most of the archaeal regulators that that can be identified on the basis of homology belong to bacterial families of regulators. In Table II the most common bacterial families of regulators are listed, and the number of paralogues is given for all archaeal genomes sequenced to date, as identified using BLAST. It is evident that of all listed protein families, members of the Lrp family are most abundantly present in archaea (for a more detailed description of Lrp-like proteins, see Chapter 3). Among the bacterial-like regulators of archaea, members of the Lrp family are unique in that they are present in all archaeal genomes to date. Like bacteria, most archaea possess multiple Lrp paralogues. Remarkably, in *F.acidarmanus* almost all of the present bacterial-like regulators are Lrp-like proteins. Most archaeal genomes contain homologues of the metal-responsive repressor proteins ArsR, FUR, and DtxR. Whereas ArsR proteins bind DNA only in the absence of their metal ion ligands, FUR and DtxR proteins require metal ions for their DNA-binding activity (5, 10, 46, 116).

In Table II the few eukaryal-like regulators present in archaea are listed as well. Multi-protein bridging factor (MBF1) is an eukaryal-like regulator that is present in all archaeal genomes sequenced to date. We have recently cloned, overproduced, and purified the *S.solfataricus* homologue of MBF1. Preliminary analysis showed that *S.solfataricus* MBF1 interacts with TBP, and the conserved archaeal genomic context of the *mbf1* gene suggests that MBF1 may be involved in regulation of proteasome activity (72, 130). Sir2 is a

| | Bacterial regulator families | | | | | | | | | | | | | | | | Eukaryal | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AraC | ArsR | Lrp | CRP/FNR | LysR | DeoR | DtxR | FUR | GntR | IclR | LacI | LuxR | MarR | MerR | Xre | MetJ | MBF1 | Sir2 |
| *Sulfolobus solfataricus* | - | 1 | 6 | - | - | - | 2 | 1 | 1 | - | - | - | 1 | - | - | - | 1 | 1 |
| *Sulfolobus tokodaii* | - | 1 | 6 | - | - | - | 2 | 1 | 1 | - | - | - | 1 | - | - | - | 1 | 1 |
| *Aeropyrum pernix* | - | 1 | 1 | - | - | - | 1 | 1 | - | - | - | - | - | - | - | - | 1 | 1 |
| *Pyrobaculum aerophilum* | - | - | 1 | - | - | - | - | 1 | - | - | - | - | - | - | - | - | 1 | 1 |
| *Pyrococcus furiosus* | - | 2 | 9 | - | - | - | 1 | - | - | - | - | - | - | - | - | - | 1 | 1 |
| *Pyrococcus horikoshii* | - | 1 | 8 | - | - | - | 1 | - | - | - | - | - | - | - | - | - | 1 | 1 |
| *Pyrococcus abyssi* | - | 2 | 7 | - | - | - | 1 | - | - | - | - | - | - | - | - | - | 1 | 1 |
| *Methanococcus jannaschii* | - | 1 | 2 | - | 1 | - | 1 | - | - | - | - | - | - | - | - | - | 1 | - |
| *Methanobacterium thermoautotrophicum* | - | 2 | 1 | - | - | - | 2 | - | - | - | - | - | - | - | - | - | 1 | - |
| *Methanosarcina barkeri* | - | 1 | 1 | - | - | - | 1 | 1 | - | - | - | - | - | - | - | - | 1 | - |
| *Archaeoglobus fulgidus* | - | 1 | 8 | - | 1 | - | 4 | 1 | - | - | - | - | - | 1* | - | - | 1 | 1 |
| *Halobacterium sp. NRC-1* | - | - | 5 | - | - | - | 3 | - | - | 1 | - | - | - | - | - | - | 2 | - |
| *Thermoplasma acidophilum* | - | - | 3 | - | - | - | 1 | 1 | 1 | - | - | - | - | - | - | - | 1 | - |
| *Thermoplasma volcanium* | - | - | 4 | - | - | - | 1 | 1 | - | - | - | - | - | - | - | - | 1 | - |
| *Ferroplasma acidarmanus* | - | - | 4 | - | - | - | 1 | - | - | - | - | - | - | - | - | - | 1 | - |
| *Bacillus subtilis* | 3 | 2 | 5 | 1 | 16 | 4 | - | 3 | 2 | 1 | 9 | - | 7 | 2 | 5 | - | - | 1 |
| *Escherichia coli* | 3 | 1 | 3 | 2 | 20 | 11 | - | 2 | 3 | 8 | 14 | 1 | 7 | 4 | - | 1 | - | - |
| *Saccharomyces cerevisiae* | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 1 | 5 |
| *Schizosaccharomyces pombe* | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 1 | 3 |

**Table II.** Number of putative archaeal regulators belonging to bacterial families of regulators. Putative proteins were assigned to one of the listed bacterial protein families on the basis of homology with bacterial family members, provided that there is homology over the entire length of the protein, with an overall expect value of $\leq 10^{-3}$ using BLAST search analysis (1). BLAST analysis and SMART analysis was used for the identification of archaeal putative regulators (http://www-archbac.u-psud.fr; http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/genom_table_cgi; http://smart.embl-heidelberg.de/). *Several archaeal genomes contain ORFs that are annotated as MerR-like regulators, but except one ORF in *A.fulgidus*, these ORFs are more related to ArsR regulators than MerR regulators. For comparison, the numbers of present bacterial regulators are also given for the bacteria *E.coli* and *B.subtilis*, as well as for the eukarya *S.cerevisiae* and *S.pombe*.

deacetylase that in eukarya modulates chromatin-mediated transcriptional regulation through post-translational modification of histones. Recently, it was shown that this protein plays a similar role in *S.solfataricus* (see below) (9). Although a homologue of Sir2 is apparently also present in *B.subtilis*, its role here is unknown.

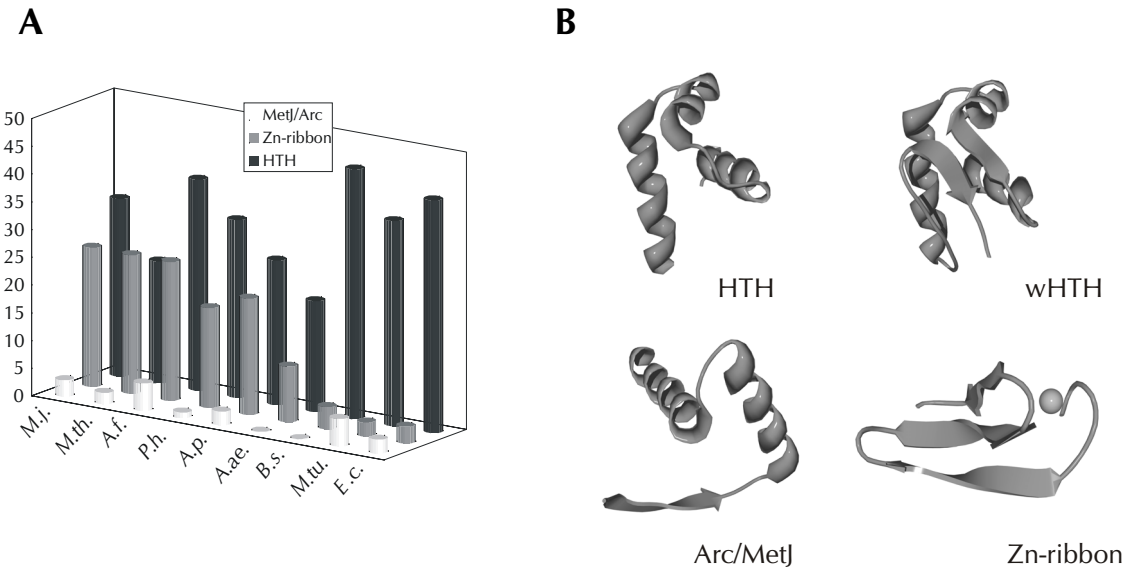



**Figure 7.** (**A**), Relative abundance of HTH, Arc/MetJ, and Zn-ribbon domains in archaea and bacteria, according Aravind & Koonin (2). Vertical axis: number of detected domains per 1000 proteins. Bacteria: *Aquifex aeolicus* (*A.ae), Bacillus subtilis* (*B.s.*), *Mycobacterium tuberculosis* (*M.tu.*); archaeal species: *Methanococcus jannaschii* (*M.j.*), *Methanobacterium thermoautotrophicum* (*M.th.*), *Archaeoglobus fulgidus* (*A.f.*), *Pyrococcus horikoshii* (*P.h.*), *Aeropyrum pernix* (*A.p.*). (**B**), Three-dimensional structure of the analyzed DNA-binding domains. HTH, helix-turn-helix (LrpA) (84); wHTH, winged-helix-turn-helix (BirA) (20); Arc/MetJ (Arc repressor) (108); Zn-ribbon (*P.furiosus* N-terminal domain of TFB) (142).

In addition to the bacterial and eukaryal regulators listed in Table II, it can be expected that archaea contain novel or archaeal-specific regulators. To screen archaeal genomes for such proteins profile search methods have been employed that identify open reading frames (ORFs) containing DNA-binding domains (2). Several conclusions could be drawn from this study. First, all archaeal genomes encode a large number of proteins containing a HTH DNA-binding domain, the majority of which belongs to the subclass of winged-HTH domains (wHTH, Fig. 7A and B). In addition to the three helices of the HTH, the wHTH contains C-terminal β-hairpin (the wing). The sequence of these HTH domains is more closely related to bacterial than to eukaryal HTH domains like POU and homeodomains. Second, the number and diversity of archaeal HTH domains is comparable to that of bacteria, and it is anticipated that the majority of HTH-containing proteins are gene or operon specific regulators. Finally, archaea contain unexpectedly large numbers of

proteins containing Arc/MetJ and Zn-ribbon domains, presumably involved in DNA binding.

In conclusion, analysis of archaeal genomes shows that they contain many members belonging to bacterial regulatory families. In particular, Lrp-like proteins are most abundantly present. In addition, most of the (novel) proteins expectedly involved in DNA-binding and transcriptional regulation contain a HTH DNA-binding domain, although proteins containing MetJ/Arc and Zn-ribbon domains are predicted as well.

## *TRANSCRIPTIONAL REGULATORY SYSTEMS IN ARCHAEA*

Although there is relatively little information yet about transcription regulation in archaea, some model systems have been studied in the recent decade. Especially for mesophilic archaea genetic tools have become available, allowing *in vivo* analysis of regulatory mechanisms. For example, heat shock response, gas vesicle formation, and bacteriorhodopsin expression have served as model systems in halophilic archaea. Functional genetic tools are not yet available for hyperthermophilic archaea, and studies focusing on these organisms have been based mainly on *in vitro* molecular analyses and comparative genomics (49). In particular, DNA-protein interaction studies have contributed considerably to an understanding of archaeal transcription regulation mechanisms. The most important archaeal regulatory systems are listed in Table II and briefly discussed below.

### Heat shock response

Using a plasmid-based *in vivo* reporter system, the regulation of two *Haloferax volcanii* heat shock-induced promoters, $P_{cct1}$ and $P_{cct2}$, has been studied (78). For basal transcription and regulation of $P_{cct1}$ it was found that sequences overlapping and immediately surrounding the TATA-element are important. The 5' upstream boundary for a fully functional and regulated $P_{cct1}$ mapped at position –37 relative to the transcriptional start site (127), presently known to coincide with the 5' boundary of the BRE (17, 86). Thus, apart from the TATA-box, no obvious *cis*-acting heat shock responsive element could be identified. It is therefore possible that the mechanism of activation is independent on direct DNA-activator interactions and relies on alternative mechanisms involving components of the basal transcription machinery, most likely TBP and/or TFB. Interestingly, it has been demonstrated that a paralogue of TFB in *H.volcanii*, called TFB2, is induced under heat

shock conditions (128). Perhaps this alternative TFB could specifically be involved in the transcription of heat shock promoters like $P_{cct1}$ and $P_{cct2}$, a mechanism that would be functionally analogous to the bacterial heat shock response, involving an alternative sigma factor. Moreover, mechanisms like these can also be expected for *Halobacterium sp.* NRC-1,


which contains at least five genes encoding TBP paralogues and seven genes encoding TFB paralogues, either on its plasmid or its chromosomal DNA (97).

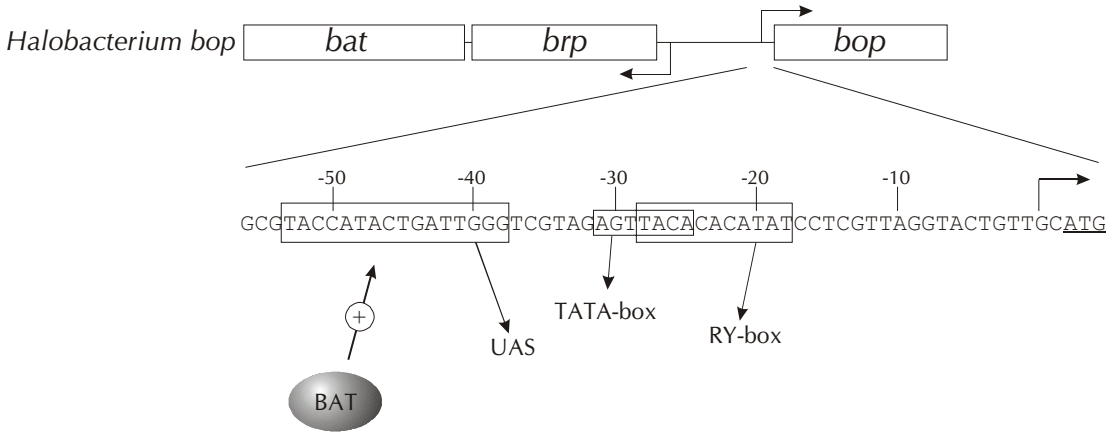
## Gas vesicle formation

Gas vesicle formation in halophilic archaea has been studied using a combination of genetic and molecular techniques. Gas vesicles are intracellular, microbial flotation devices that consist of mainly one structural protein (GvpA), although a total of 14 genes are involved in the formation of gas-vesicles. All these genes are organized in two divergently transcribed clusters, *gvpDEFGHIJKLM*, and *gvpACNO*, with two divergent promoters, $P_{gvpD}$ and $P_{gvpA}$, respectively (68). Expression of chromosomally encoded gas vesicle genes is regulated in response to salt concentration and growth phase (42, 43, 113). The two proteins GvpD and GvpE appear to play a key role in this regulation. GvpD apparently acts as a repressor since its deletion leads to overproduction of gas vesicles (44). In contrast, GvpE acts as an activator for the $P_{gvpA}$ promoter (53, 76), most likely by binding to an inverted repeat element upstream of the $P_{gvpA}$ promoter. The GvpE protein has been analyzed using molecular modeling, and was proposed to contain a leucine-zipper domain reminiscent of eukaryal bZIP transcription factors like yeast GCN4 (71, 76).

## Expression of bacteriorhodopsin

Bacteriorhodopsin is the main component of the 'purple membrane' of *Halobacterium* species, and acts as a light-driven proton pump generating a membrane potential that is used to drive ATP synthesis (75). Bacteriorhodopsin consists of retinal complexed with bacterio-opsin, and synthesis of the latter is encoded by the *bop* gene. Regulation of *bop* transcription is responsive to oxygen and light (118), and has been studied using a *Halobacterium* sp. NRC-1 *in vivo* system. Two genes, *brp* and *bat*, transcribed divergently from the *bop* promoter, are necessary for *bop* transcription (85). Whereas *brp* is thought to be

involved in the biosynthesis of retinal, the *bat* gene product has been shown to be a regulator for *bop* transcription (54). Using saturation mutagenesis, *cis*-elements in the *bop* promoter were identified. While wild-type promoter activity was found to reside in the 53-bp region upstream from the transcriptional start (55) the importance of the TATA-box and an upstream activating sequence (UAS) for *bop* transcription was demonstrated (7). Subsequently, the Bat protein was identified as the *trans*-acting factor of the UAS, not only for *bop*, but also for several other genes involved in bacteriorhodopsin synthesis containing a similar UAS in their promoters (8). Most likely, Bat is an HTH-containing regulator for




the bacteriorhodopsin regulon, capable of responding to light and oxygen through a putative photo-responsive GAF domain and a redox-responsive PAS/PAC domain (8). GAF domains are reminiscent of phytochromes, and exist in two forms that are reversibly interconvertible by light (3), whereas PAS domains detect their signal by way of an associated cofactor such as heme, flavin, or 4-hydroxycinnamyl chromophore (143). PAC domains are frequently observed C-terminally of PAS domains, and are proposed to contribute to the PAS domain fold. An interesting aspect of regulation of the *bop* promoter is its sensitivity to DNA supercoiling. Novobiocin, a DNA gyrase inhibitor reduces *bop* transcription up to 10-fold, at concentrations subinhibitory for growth. An 11-bp alternating purine-pyrimidine element (R-Y box) centered immediately downstream of the TATA-box was shown to adopt a non-B-DNA conformation under inducing conditions (high negative supercoiling) (140), and mutations in this R-Y-box tempered the effect of DNA supercoiling (6). Furthermore, maximal *bop* promoter activity could not be related to the presence of a canonical BRE. Since the genome of *Halobacterium* NRC-1 encodes multiple TBP and TFB paralogues, it is possible that *bop* transcription involves such an alternative general transcription factor, as seems to be the case in *H.volcanii* heat shock response (6, 97, 128) (see above). Regulation of *bop* transcription is thus a complex system

involving multiple regulatory factors, and the availability of genetic tools for *Halobacterium* makes it an attractive model system to study archaeal transcription regulation *in vivo*.

## Nitrogen fixation

In several methanogenic archaea the expression of genes involved in nitrogen fixation has been studied. In *Methanococcus maripaludis* a palindromic repressor-binding site was identified immediately downstream the transcriptional start site of the nitrogen-regulated *nifH* promoter. Mutation of this site caused marked derepression of transcription, as determined using an *in vivo* $P_{nifH}$-*lacZ* reporter system (32). Subsequently, an unidentified factor in *M.maripaludis* extracts was shown to bind to this site. A similar repressor-binding site was identified in the nitrogen-regulated *M.maripaludis glnA* promoter, in this case just upstream the transcriptional start (33). In *Methanosarcina barkeri*, a different mechanism appears to be employed. When ammonia is present, an unidentified negatively acting factor appears to prevent binding of one of the general transcription factors (30). This unidentified factor itself is apparently not involved in DNA-binding, and may act indirectly through other regulatory factors.

## Carbon source utilization

The regulation of several genes involved in carbon source utilization has been studied in hyperthermophilic archaea. For instance, several *P.furiosus* maltose-regulated genes were identified on the basis of their induction during growth on maltose (111, 112). Although the function of several of these genes is unknown, some of these were induced up to 100-fold in response to maltose. In addition, DNA microarrays were used to determine the effect of elemental sulfur on maltose-grown *P.furiosus* cells (117). 21 ORFs, most of them encoding hydrogenase systems, were down-regulated more than five-fold, whereas nine ORFs were upregulated at least six-fold.

In the same organism, coordinated induction of divergently oriented genes by β-linked glucose polymers has been demonstrated (132). The locus contains *celB* (encoding an intracellular β-glucosidase), *lamA* (encoding an extracellular β-1.3-endoglucanase), and *adhAB* (encoding two alcohol dehydrogenases). Although mRNA encoding these genes is abundantly present after induction, *in vitro* transcription experiments showed that both divergently oriented promoters are intrinsically weak (132). The presence of number of(inverted) repeat elements present in the intergenic region suggests that binding of an activator is required for efficient transcription.

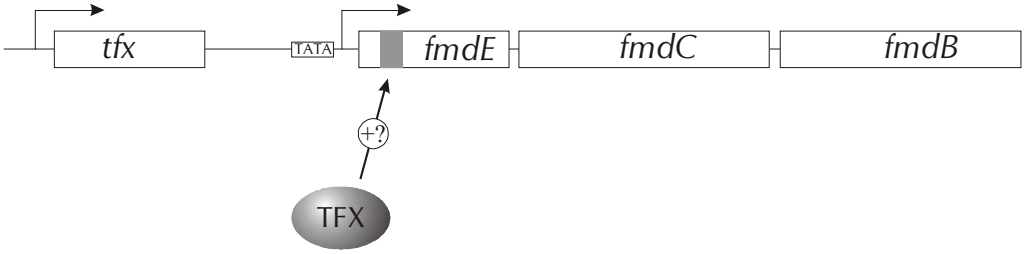
A comparison of promoter sequences of *P.furiosus* glycolytic genes revealed the presence of a common ATCACN$_5$GTGAT inverted repeat (131). This inverted repeat is not only present in the promoters of glycolysis genes, but also in the promoters of other sugar metabolism genes, e.g. encoding intracellular or extracellular α-amylases. It is hypothesized that this inverted repeat is a regulatory element that is involved in the coordinated regulation of these genes. The putative trans-acting factor involved in this coordinated regulation has not been identified yet, but as most of the inverted repeats are located downstream of the TATA-boxes of the respective promoters, it seems likely that transcription is negatively controlled.

Another example of coordinated regulation of carbon-utilization genes has been described for *S.solfataricus*. However, in contrast to the situation described above, the involved genes are not physically linked on the *S.solfataricus* genome. It was shown that the expression of *malA*, (encoding an α-glucosidase), *lacS* (encoding a β-glycosidase), and an α-amylase was induced when supplementary carbon sources were removed from a defined sucrose medium (60). The effect was most prominent when yeast extract was the supplementary carbon source. Conversely, when yeast extract was added to the defined sucrose medium, the expression of the same genes was inhibited.

### Tfx

The gene encoding Tfx was identified directly upstream from the *fmdECB* operon of *M.thermoautotrophicum*, which encodes a molybdenum formylmethanofuran dehydrogenase.

*M.thermoautotrophicum fmdECB*



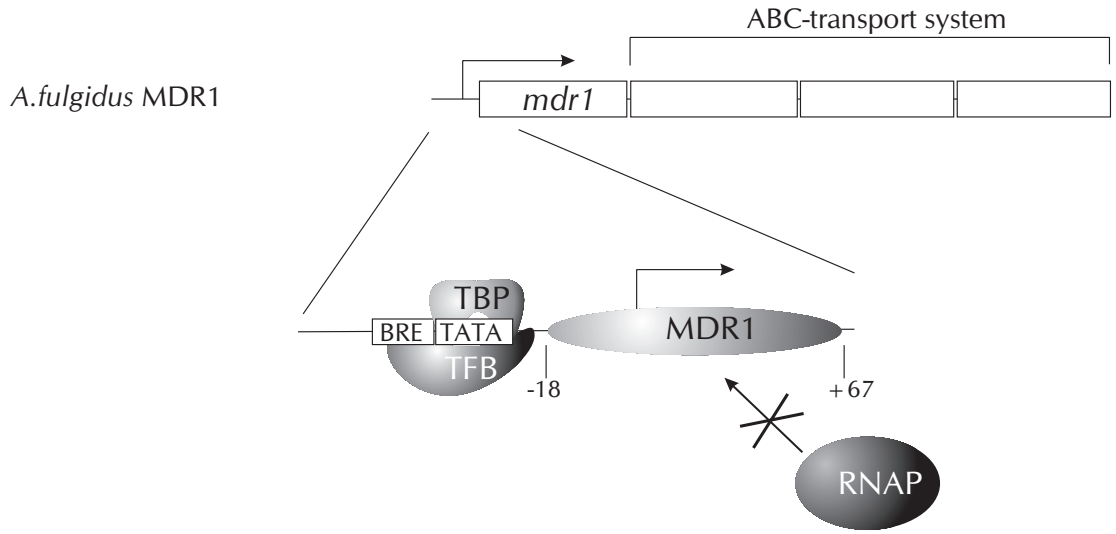

Primary sequence analysis revealed that although the protein has no obvious homologues in the database, it contains a putative HTH DNA-binding domain. Tfx binds to a site located 167 bp downstream of the transcriptional start site of *fmdE*, and it was proposed that Tfx is a transcriptional activator required for the expression of *fmdECB* (67).

The gene encoding a homologue of Tfx in *Pyrococcus* species is clustered with *iorAB* and *acsIIA* genes, encoding proteins involved in the fermentation of aromatic amino acids.

Although experiments in our laboratory showed that the purified recombinant Tfx homologue bound efficiently to the *iorAB-acsIIA* promoter, the specificity of binding was rather low, since binding of the *Pyrococcal* TFx homologue to unrelated control promoters occurred with comparable affinity (21).

## MDR1

MDR1 is the first example of an archaeal regulator from which the mechanism of transcriptional repression has been characterized in detail, both *in vitro* and *in vivo* (10). MDR1 is an *Archaeoglobus fulgidus* homologue of the bacterial metal-dependent transcriptional repressor DtxR. The *mdr1* gene is clustered with three genes that putatively encode an iron-importing ABC transport system. Transcription from the *mdr1* promoter results in a polycistronic mRNA containing *mdr1* and the three downstream ABC transporter-encoding genes. In addition, a monocistronic *mdr1* mRNA is produced. It was

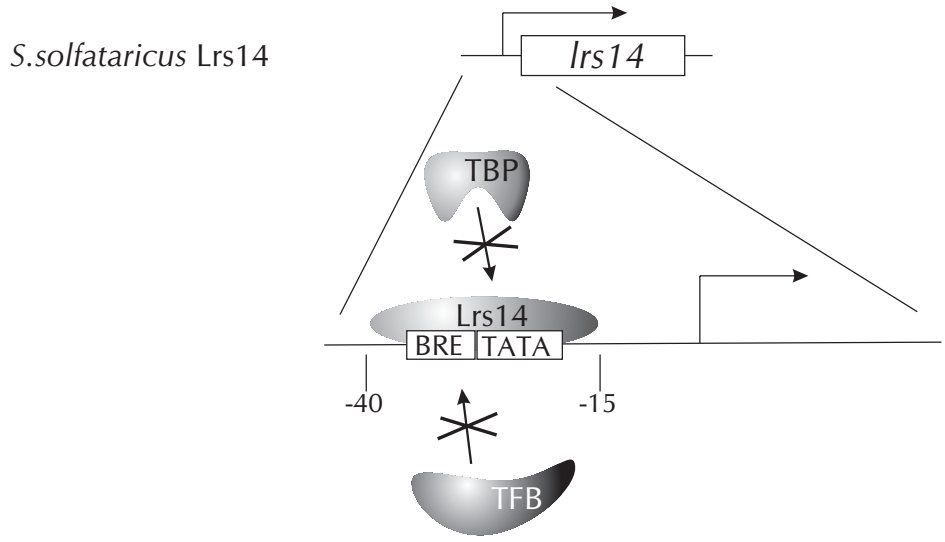



found that transcription from the *mdr1* promoter was induced after depletion of metal ions in the medium. Subsequent experiments showed that the MDR1 protein specifically binds and represses the cluster's promoter in the presence of $Fe^{2+}$, $Mn^{2+}$, or $Ni^{2+}$, not only *in vitro* but also *in vivo*. Using a variety of footprinting techniques and mutational analysis of MDR1 binding sites, MDR1 was shown to bind cooperatively to three adjacent operators in the promoter region, ranging from –18 to +67 relative to the transcriptional start site. Binding of MDR1 to this region did not abolish formation of the DNA-TBP-TFB complex; instead, simultaneous binding of MDR1, TBP, and TFB to the DNA was demonstrated.

Binding of MDR1 did however prevent RNAP recruitment by acting as a physical block to RNAP binding (10).

### Lrs14

Although Lrs14 from *S.solfataricus* was initially annotated as a member of the Lrp family of regulators (96), a more thorough inspection of its amino acid sequence revealed that it is rather related to a class of DNA-binding proteins so far only encountered in (cren)archaea (see Chapter 3). Lrs14 is mainly expressed in exponentially growing *S.solfataricus* cells as a monocistronic mRNA (96). Although its physiological target gene(s) have not been identified yet, Lrs14 has been successfully used to study the mechanism of negative autoregulation in archaea. Lrs14 was shown to bind the *lrs14* promoter and repress transcription in the absence of any ligand (12, 96). In contrast to MDR1, which prevents RNAP recruitment by binding downstream from the TATA-box, Lrs14 binds to a DNA




sequence from position –60 to +1 relative to the transcriptional start. While Lrs14 recognizes multiple binding sites within this region, it specifically recognizes the bases -30 to –22, which also comprise the TATA-box. In agreement with these results, it was shown the Lrs14 and the TBP/TFB complex compete for the same binding site, and that binding of Lrs14 and TBP/TFB are indeed mutually exclusive events: either one of the protein-DNA complexes is stable once established (12). Instead of blocking RNAP recruitment, Lrs14 thus represses transcription by blocking TBP/TFB binding.

## SaLrp

As a member of the Lrp-family of regulators (see Chapter 3), SaLrp was identified in the genome of *S.acidocaldarius*. SaLrp is expressed predominantly during the stationary phase as a monocistronic transcript, and binds to its own promoter with relatively weak specificity, since binding could also be detected to the *E.coli lrp* and *P.furiosus lrpA* promoters. SaLrp-DNA binding appears to be leucine-independent (45).

## Repression by chromatin

A typical feature of eukaryal transcription is that free assembly of the PIC cannot take place since nucleosomes prevent the binding of TBP to the TATA-box. Activation of transcription therefore involves chromatin modification or remodeling, in many cases initiated by the binding of an activator which access to the DNA is permitted. Thus, in striking contrast to the situation in bacteria, a strong eukaryal core promoter is essentially inactive because of this repressive ground state, and mechanisms of regulation are therefore fundamentally different in the bacterial and eukaryal domain (125).

Recently, it was shown that ALBA (Acetylation Lowers Binding Affinity), one of the chromatin-associated proteins of *S.solfataricus* (previously named Sso10), is acetylated *in vivo*, and that this acetylation antagonizes the repressive capacity of the protein by decreasing its DNA-binding affinity. Sir2, a homologue of the eukaryal Sir2, interacts with and deacetylates ALBA, thereby increasing its DNA-binding affinity and its repressive effect on transcription (9). This clearly shows that eukaryal-like mechanisms of transcriptional regulation also exist in archaea. Future studies are required to determine whether the role of regulation by such chromatin-associated proteins is comparable to the role of eukaryal chromatin in transcriptional regulation. It is tempting to assume that the complete range of archaeal transcriptional regulatory mechanisms is a mosaic of bacterial and eukaryal features.

| Organism | Cellular process | Genes regulated | Identified regulator | Regulatory signal | References |
|---|---|---|---|---|---|
| *Hf.volcanii* | heat-shock response | *cct1*, *cct2* | ? | heat shock | (78, 127, 128) |
| *Hb.salinarium*, *Hf.mediterranei* | gas-vesicle formation | *gvpACNO*, *gvpDEFGHIJKLM* | GvpD, GvpE | salt concentration, growth phase | (42-44, 53, 76, 113) |
| *Halobacterium* NRC-1 | bacteriorhodopsin synthesis | *bop*, and related genes | Brp, Bat | oxygen, light | (6, 8, 54, 55, 118) |
| *M.maripaludis* | nitrogen fixation | *nifH*, *glnA* | ? | ammonia | (32, 33) |
| *P.furiosus* | carbon source utilization | glycolytic and other | ? | maltose | (111, 112) |
| *P.furiosus* | carbon source utilization | *celB*, *lamA*, *adhAB* | ? | β-linked glucose polymers | (132) |
| *S.solfataricus* | carbon source utilization | *malA*, *lacS* | ? | ? (yeast extract) | (60) |
| *M.thermoautotrophicum* | methanogenesis | *fmdECB* | Tfx | molybdenum, tungsten | (67) |
| *A.fulgidus* | ABC transport system | *mdr1*, ABC-transport genes | Mdr1 | $Fe^{2+}$, $Mn^{2+}$, $Ni^{2+}$ | (10) |
| *S.solfataricus* | unknown | *lrs14* | Lrs14 | ? | (12, 96) |
| *S.acidocaldarius* | unknown | ? | SaLrp | ? | (45) |
| *S.solfataricus* | unknown | global (?) | Alba (chromatin) | ? | (9) |

**Table III.** Overview of studied archaeal transcriptional regulatory systems

## CONCLUDING REMARKS

Transcriptional regulation in archaea has been studied using a number of model systems, but many of the responsible *cis* and *trans*-acting factors still remain to be identified. The regulatory proteins that have been characterized in more detail are either homologues of bacterial regulators (MDR1, SaLrp), archaeal-specific regulators with bacterial-like DNA-binding domains (Bat, Tfx, Lrs14), or archaeal-specific with eukaryal-like DNA-binding domains (GvpE). Only few homologues of eukaryal regulators can be identified in archaeal genomes, and except for Sir2, detailed characterization of these proteins is still lacking.

The fact that archaeal transcription involves eukaryal-like and bacterial-like features evokes speculation on the evolution of transcriptional machineries and their regulators in the three domains of life. With the current knowledge the strict terms "bacterial-like regulators" (B) and "eukaryal-like basal transcription machinery" (E) are rather confusing, and it would be more appropriate to refine these terms as "bacterial/archaeal regulators" (BA) and archaeal/eukaryal transcriptional machinery" (AE) (11). A model has been described in which the development of basal transcription-regulation relationships within the three domains of life is fitted onto the evolutionary tree of life based on rRNA sequences (see Fig. 8) (11). In this model it is proposed that the last common ancestor of all current life possessed BA-type regulators, since it is most likely that these must have been present before the bacterial lineage diverged from the archaeal/eukaryal lineage (11). The identity of the ancestral basal transcription machinery cannot be ascertained, and could have been of B-type or AE-type (note that all

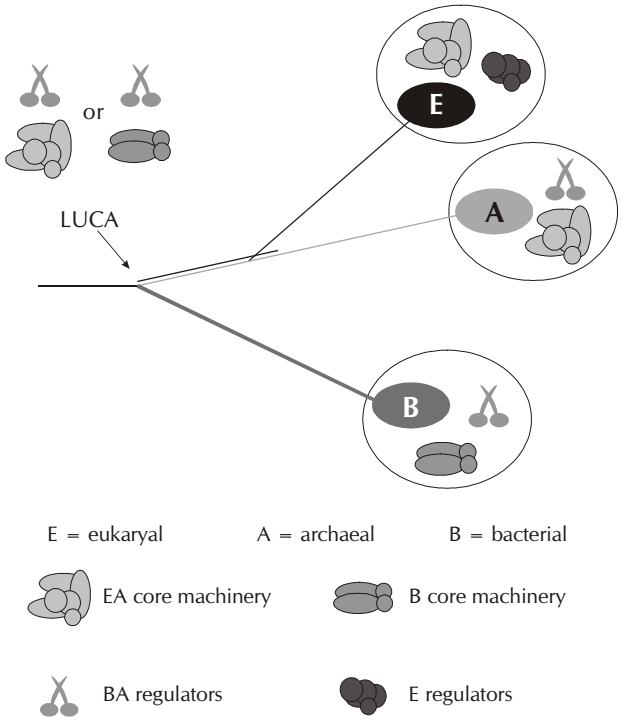



**Figure 8.** Model for evolution of relationships between basal transcription machinery's and regulators, according Bell & Jackson (11). LUCA, last common ancestor of all life. See text for further explanation.

bacterial RNAP subunits except the σ-factor share homology with AE-type RNAP subunits, Table I). In the first case the B-type RNAP gained more complexity in the lineage that gave rise to the present-day archaea and eukarya. In the second case, the AE-type RNAP lost its complexity in the bacterial lineage after this lineage diverged. BA-type regulators were retained in the bacteria and archaea, but not in the eukarya. Possibly, the increase in genome size of the latter caused the development of a highly efficient histone-dependent DNA compaction system that was subsequently integrated in transcriptional regulatory processes, and eventually led to a more complex and fundamentally different mode of transcriptional regulation and the concomitant loss of BA-type regulators (125).

## *AIMS AND OUTLINE OF THE THESIS*

The aim of the research presented here was to gain insight in the mechanisms by which transcription in hyperthermophilic archaea is regulated. To accomplish this, we have aimed (I) to identify transcriptional regulatory proteins from hyperthermophilic archaea, (II) to characterize these proteins, and (III) to determine how these proteins modulate the process of transcription initiation.

**Chapter 1** describes the characteristics of the archaeal transcription machinery, and compiles transcription-related data that was obtained in the past two decades. The archaeal transcription machinery appears to be a simplified version of the eukaryal RNA polymerase II system, lacking various general transcription factors that are essential for eukaryal transcription initiation. However, archaeal genomes encode TFE, a homologue eukaryal TFIIEα transcription factor. Its stimulatory role in transcription is described in **Chapter 2**.

Although the archaeal transcription machinery is eukaryal-like, many genes encoding members of bacterial regulatory protein families can be found within archaeal genomes. Members of the Lrp family are most abundantly present in archaea, and **Chapter 3** describes the properties of Lrp-like proteins. When this research project was started, fully sequenced archaeal genomes just became available (22). Only the gene encoding LrpA from *P.furiosus* had previously been identified in our laboratory and by others (79, 83), and our initial strategy included the characterization of this protein, which is described in **Chapter 4**. LrpA was shown to negatively autoregulate its own transcription in a ligand-independent manner. The efficient production and purification of recombinant LrpA enabled crystallization of the protein and **Chapter 5** describes its resolved three-dimensional structure, which is the first structure of a member of the Lrp family. Subsequently, during

the participation of our laboratory in the *S.solfataricus* P2 genome sequencing project, we were able to identify candidate regulatory genes in a more directed, bioinformatics-based approach, resulting in the identification and characterization of LysM and ChoR. LysM is another example of an archaeal Lrp-like protein, and in **Chapter 6** we have used the genomic context of LysM in the *S.solfataricus* genome to experimentally identify its physiological target and ligand. This study indicates for the first time that an Lrp-like protein may activate archaeal transcription.

Besides bacterial-like regulators, archaeal genomes encode unique archaeal-specific regulators that can be identified on the basis of a present DNA-binding domain. A putative regulator for copper homeostasis (ChoR) was identified in the *S.solfataricus* genome on the basis of a predicted HTH DNA-binding domain and a metal-binding domain. In **Chapter 7** it is demonstrated that ChoR is indeed a metal-responsive DNA-binding protein that is most likely involved in the repression of a heavy metal-efflux system.

**Chapter 8** summarizes the data presented in this thesis, and adds some concluding remarks with respect to the implications of the work.

## REFERENCES


1. Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-402.
2. Aravind, L., Koonin, E.V. 1999. DNA-binding proteins and evolution of transcription regulation in the archaea. Nucleic Acids Res 27:4658-4670.
3. Aravind, L., and C.P. Ponting 1997. The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem Sci 22:458-9.
4. Arents, G., and E.N. Moudrianakis 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc Natl Acad Sci U S A 92:11170-4.
5. Bairoch, A. 1993. A possible mechanism for metal-ion induced DNA-protein dissociation in a family of prokaryotic transcriptional regulators. Nucleic Acids Res 21:2515.
6. Baliga, N.S., and S. Dassarma 2000. Saturation mutagenesis of the haloarchaeal *bop* gene promoter: identification of DNA supercoiling sensitivity sites and absence of TFB recognition element and UAS enhancer activity. Mol Microbiol 36:1175-83.
7. Baliga, N.S., and S. DasSarma 1999. Saturation mutagenesis of the TATA box and upstream activator sequence in the haloarchaeal *bop* gene promoter. J Bacteriol 181:2513-8.
8. Baliga, N.S., S.P. Kennedy, W.V. Ng, L. Hood, and S. DasSarma 2001. Genomic and genetic dissection of an archaeal regulon. Proc Natl Acad Sci U S A 98:2521-5.
9. Bell, S.D., C.H. Botting, B.N. Wardleworth, S.P. Jackson, and M.F. White 2002. Alba, a conserved archaeal chromatin protein, interacts with Sir2 and is regulated by acetylation. Science. *In press.*
10. Bell, S.D., S.S. Cairns, R.L. Robson, S.P. Jackson 1999. Transcriptional regulation of an archaeal operon *in vivo* and *in vitro*. Mol Cell 4:971-982.
11. Bell, S.D., and S.P. Jackson 2001. Mechanism and regulation of transcription in archaea. Curr Opin Microbiol 4:208-13.
12. Bell, S.D., and S.P. Jackson 2000. Mechanism of autoregulation by an archaeal transcriptional repressor. J Biol Chem 275:31624-9.
13. Bell, S.D., and S.P. Jackson 2000. The role of transcription factor B in transcription initiation and promoter clearance in the archaeon *Sulfolobus acidocaldarius*. J Biol Chem 275:12934-40.
14. Bell, S.D., and S.P. Jackson 1998. Transcription in Archaea, p. 41-51, Mechanisms of transcription, vol. LXIII. Cold Spring Harbor Laboratory Press.
15. Bell, S.D., S.P. Jackson 1998. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends in Microbiology 6:222-228.
16. Bell, S.D., C. Jaxel, M. Nadal, P.F. Kosa, and S.P. Jackson 1998. Temperature, template topology, and factor requirements of archaeal transcription. Proc Natl Acad Sci U S A 95:15218-22.
17. Bell, S.D., P.L. Kosa, P.B. Sigler, S.P. Jackson 1999. Orientation of the transcription preinitiation complex in archaea. Proc Natl Acad Sci U S A 96:13662-13667.
18. Best, A.A., and G.J. Olsen 2001. Similar subunit architecture of archaeal and eukaryal RNA polymerases. FEMS Microbiol Lett 195:85-90.
19. Breathnach, R., and P. Chambon 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem 50:349-83.

20. Brennan, R.G., S. Vasu, B.W. Matthews, and A.J. Otsuka 1989. Crystallization of the bifunctional biotin operon repressor. J Biol Chem 264:5.
21. Brinkman, A.B., and J. van der Oost, *Unpublished data.*
22. 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. Geoghagen, and J.C. Venter 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii.* Science 273:1058-1073.
23. Buratowski, S., and H. Zhou 1993. Functional domains of t ranscription factor TFIIB. Proc Natl Acad Sci U S A 90:5633-7.
24. Burgess, R.R. 1969. Separation and characterization of the subunits of ribonucleic acid polymerase. J Biol Chem 244:6168-76.
25. Burke, T.W., P.J. Willy, A.K. Kutach, J.E.F. Butler, and J.T . Kadonaga 1998. The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box, p. 75 -82, Mechanisms of transcription, vol. LXIII. Cold Spring Harbor Laboratory Press.
26. Charbonnier, F., and P. Forterre 1994. Compar ison of plasmid DNA topology among mesophilic and thermophilic eubacteria and archaeabacteria. J Bacteriol 176:1251 -9.
27. Chartier, F., B. Laine, D. Belaiche, and P. Sautiere 1989. Primary structure of the chromosomal proteins MC1a, MC1b, and MC1c from th e archaebacterium *Methanothrix soehngenii.* J Biol Chem 264:17006-15.
28. Chartier, F., B. Laine, D. Belaiche, J.P. Touzel, and P. Sautiere 1989. Primary structure of the chromosomal protein MC1 from the archaebacterium *Methanosarcina* sp. CHTI 55. Biochim B iophys Acta 1008:309-14.
29. Chartier, F., B. Laine, and P. Sautiere 1988. Characterization of the chromosomal protein MC1 from the thermophilic archaebacterium *Methanosarcina* sp. CHTI 55 and its effect on the thermal stability of DNA. Biochim Biophys Acta 951:149-56.
30. Chien, Y., J.D. Helmann, and S.H. Zinder 1998. Interactions between the promoter regions of nitrogenase structural genes (*nifHDK2*) and DNA-binding proteins from $N_2$- and ammonium-grown cells of the archaeon *Methanosarcina barkeri* 227. J Bacteriol 180:2723-8.
31. Choli, T., P. Henning, B. Wittmann -Liebold, and R. Reinhardt 1988. Isolation, characterization and microsequence analysis of a small basic methylated DNA -binding protein from the Archaebacterium, *Sulfolobus solfataricus.* Biochim Biophys Acta 950:193-203.
32. Cohen-Kupiec, R., C. Blank, J.A. Leigh, 1997. Transcriptional regulation in Archaea: *in vivo* demonstration of a repressor binding site in a methanogen. Proc Natl Acad Sci U S A 94:1316 -1320.
33. Cohen-Kupiec, R., C.J. Marx, J.A. Leigh, 1999. Function and regulation of *glnA* in the methanogenic archaeon *Methanococcus maripaludis.* J Bacteriol 181:256-261.
34. Cox, J.M., M.M. Hayward, J.F. Sanchez, L.D. Gegnas, S. van der Zee, J.H. Dennis, P.B. Sigler, and A. Schepartz 1997. Bidirectio nal binding of the TATA box binding protein to the TATA box. Proc Natl Acad Sci U S A 94:13475-80.
35. Cramer, P., D.A. Bushnell, and R.D. Kornberg 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science 292:1863 -76.
36. Dahmus, M.E. 1995. Phosphorylation of the C -terminal domain of RNA polymerase II. Biochim Biophys Acta 1261:171-82.

37. Dalgaard, J.Z., and R.A. Garret 1993. Archaeal hyperthermophile genes. In: Kates (ed.), The biochemistry of archaea. Elsevier Scienc e Publishers B.V.
38. Danner, S., and J. Soppa 1996. Characterization of the distal promoter element of halobacteria *in vivo* using saturation mutagenesis and selection. Mol Microbiol 19:1265 -76.
39. Darcy, T.J., W. Hausner, D.E. Awery, A.M. Edwards, M. Tho mm, and J.N. Reeve 1999. *Methanobacterium thermoautotrophicum* RNA polymerase and transcription *in vitro*. J Bacteriol 181:4424 -9.
40. DeDecker, B.S., R. O'Brien, P.J. Fleming, J.H. Geiger, S.P. Jackson, and P.B. Sigler 1996. The crystal structure of a hyper thermophilic archaeal TATA -box binding protein. J Mol Biol 264:1072 -84.
41. Eloranta, J.J., A. Kato, M.S. Teng, and R.O. Weinzierl 1998. *In vitro* assembly of an archaeal D -L-N RNA polymerase subunit complex reveals a eukaryote -like structural arrangement. Nucleic Acids Res 26:5562-7.
42. Englert, C., K. Kruger, S. Offner, and F. Pfeifer 1992. Three different but related gene clusters encoding gas vesicles in halophilic archaea. J Mol Biol 227:586 -92.
43. Englert, C., and F. Pfeifer 1993. Analysis of gas ves icle gene expression in *Haloferax mediterranei* reveals that GvpA and GvpC are both gas vesicle structural proteins. J Biol Chem 268:9329 -36.
44. Englert, C., G. Wanner, and F. Pfeifer 1992. Functional analysis of the gas vesicle gene cluster of the halophilic archaeon *Haloferax mediterranei* defines the *vac*-region boundary and suggests a regulatory role for the *gvpD* gene or its product. Mol Microbiol 6:3543 -50.
45. Enoru-Eta, J., D. Gigot, T.L. Thia -Toong, N. Glansdorff, and D. Charlier 2000. Purification an d 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 182:3661-72.
46. Escolar, L., J. Perez -Martin, and V. de Lorenzo 1999. Opening the iron box: transcriptional metalloregulation by the Fur protein. J Bacteriol 181:6223 -9.
47. Fiala, G., K.O. Stetter 1986. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeabacteria growing optim ally at 100°C. Arch Microbiol 145:56 -61.
48. Frey, G., M. Thomm, B. Brüdigam, H.P. Gohl, and W. Hausner 1990. An archaebacterial cell -free transcription system. The expression of tRNA genes from *Methanococcus vannielii* is mediated by a transcription factor. Nucleic Acids Res 18:1361 -1367.
49. Gelfand, M.S., E.V. Koonin, and A.A. Mironov 2000. Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. Nucleic Acids Res 28:695 -705.
50. Gentry, D. R. 1989. *rpoZ*, encoding the omega subunit of *Escherichia coli* RNA polymerase, is in the same operon as *spoT*. J Bacteriol 171:1271 -1277.
51. Grayling, R.A., K.A. Bailey, and J.N. Reeve 1997. DNA binding and nuclease protection by the HMf histones from the hyperthermophilic archaeon *Methanothermus fervidus*. Extremophiles 1:79 -88.
52. Grayling, R.A., K. Sandman, and J.N. Reeve 1996. Histones and chromatin structure in hyperthermophilic Archaea. FEMS Microbiol Rev 18:203 -13.
53. Gregor, D., and F. Pfeifer 2001. Use of a halobacterial *bgaH* reporter gene to analyse the regulation of gene expression in halophilic archaea. Microbiology 147:1745 -54.
54. Gropp, F., and M.C. Betlach 1994. The *bat* gene of *Halobacterium halobium* encodes a trans -acting oxygen inducibility factor. Proc Natl Acad Sci U S A 91:5475-9.

55. Gropp, F., R. Gropp, and M.C. Betlach 1995. Effects of upstream deletions on light - and oxygen-regulated bacterio-opsin gene expression in *Halobacterium halobium*. Mol Microbiol 16:357-64.
56. Gross, C. A., M. Lonetto, and R. Losick 1992. Bacterial sigma factors, Transcripitonal regulation vol. 1. Cold Spring Harbor Laboratory Press.
57. Guagliardi, A., A. Napoli, M. Rossi, and M. Ciaramella 1997. Annealing of complementary DNA strands above the melting point of the duplex promoted by an ar chaeal protein. J Mol Biol 267:841-8.
58. Hamiche, A., V. Carot, M. Alilat, F. De Lucia, M.F. O'Donohue, B. Revet, and A. Prunell 1996. Interaction of the histone $(H3-H4)_2$ tetramer of the nucleosome with positively supercoiled DNA minicircles: Potential flipping of the protein from a left- to a right-handed superhelical form. Proc Natl Acad Sci U S A 93:7588-93.
59. Hamiche, A., and H. Richard-Foy 1998. The switch in the helical handedness of the histone $(H3-H4)_2$ tetramer within a nucleoprotein particle req uires a reorientation of the H3-H3 interface. J Biol Chem 273:9261-9.
60. Haseltine, C., R. Montalvo-Rodriguez, E. Bini, A. Carl, and P. Blum 1999. Coordinate transcriptional control in the hyperthermophilic archaeon *Sulfolobus solfataricus*. J Bacteriol 181:3920-7.
61. Hausner, W., G. Frey, and M. Thomm 1991. Control regions of an archaeal gene. A TATA box and an initiator element promote cell-free transcription of the $tRNA^{Val}$ gene of *Methanococcus vannieli*. J Mol Biol 222:495-508.
62. Hausner, W., U. Lange, and M. Musfeldt 2000. Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. J Biol Chem 275:12393-9.
63. Hausner, W., and M. Thomm 2001. Events during initiation of archaeal transcription: open complex formation and DNA-protein interactions. J Bacteriol 183:3025-31.
64. Heil, A., Zillig, W. 1970. Purification of bacterial DNA-dependent RNA polymerase from isolated subunits as a tool for the elucidation of the role of the subunits in transcription. FEBS Lett 11:165-168.
65. Hethke, C., A. Bergerat, W. Hausner, P. Forterre, and M. Thomm 1999. Cell-free transcription at 95 degrees: thermostability of transcriptional components and DNA topology requirements of *Pyrococcus* transcription. Genetics 152:1325-33.
66. Hethke, C., A.C. Geerling, W. Hausner, W.M. de Vos, and M. Thomm 1996. A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. Nucleic Acids Res 24:2369-2376.
67. Hochheimer, A., Hedderich, R., Thauer, R.K. 1999. The DNA binding protein Tfx f rom *Methanobacterium thermoautotrophicum*: structure, DNA binding properties and transcriptional regulation. Mol Microbiol 32:641-650.
68. Horne, M., C. Englert, C. Wimmer, and F. Pfeifer 1991. A DNA region of 9 kbp contains all genes necessary for gas vesicle synthesis in halophilic archaeabacteria. Mol Microbiol 5:1159-74.
69. Hudepohl, U., W.D. Reiter, and W. Zillig 1990. *In vitro* transcription of two rRNA genes of the archaebacterium *Sulfolobus* sp. B12 indicates a factor requirement for specific initiati on. Proc Natl Acad Sci U S A 87:5851-5.
70. Karantza, V., A.D. Baxevanis, E. Freire, and E.N. Moudrianakis 1995. Thermodynamic studies of the core histones: ionic strength and pH dependence of H2A-H2B dimer stability. Biochemistry 34:5988-96.
71. Konig, P., and T.J. Richmond 1993. The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. J Mol Biol 233:139-54.

72. Koonin, E.V., Y.I. Wolf, and L. Aravind 2001. Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. Genome Res 11:240-52.
73. Kornberg, R.D. 1999. Eukaryotic transcriptional control. Trends Cell Biol 9:M46-9.
74. Kosa, P.F., G. Ghosh, B.S. DeDecker, P.B. Sigler 1997. The 2.1-Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. Proc Natl Acad Sci U S A 94:6042-6047.
75. Krebs, M.P., and H.G. Khorana 1993. Mechanism of light-dependent proton translocation by bacteriorhodopsin. J Bacteriol 175:1555-60.
76. Krüger, K., Hermann, T., Armbruster, V., Pfeifer, F. 1998. The transcriptional activator GvpE for the halobacterial gas vesicle genes resembles a basic region leucine-zipper regulatory protein. J Mol Biol 279:761-771.
77. Kunst, F., N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo, M.G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S.C. Brignell, S. Bron, S. Brouillet, C.V. Bruschi, B. Caldwell, V. Capuano, N.M. Carter, S.K. Choi, J.J. Codani, I.F. Connerton, A. Danchin, and *et al.* 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390:249-56.
78. Kuo, Y.P., Thompson, D.K., St Jean, A., Charlebois, R.L., Daniels, C.J. 1997. Characterization of two heat shock genes from *Haloferax volcanii*: a model system for transcription regulation in the Archaea. J Bacteriol 179:6318-6324.
79. Kyrpides, N.C., Ouzounis, C.A. 1995. The eubacterial transcriptional activator Lrp is present in the archaeon *Pyrococcus furiosus*. Trends Biochem Sci 20:140-141.
80. Lagrange, T., A.N. Kapanidis, H. Tang, D. Reinberg, and R.H. Ebright 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev 12:34-44.
81. Lalo, D., C. Carles, A. Sentenac, and P. Thuriaux 1993. Interactions between three common subunits of yeast RNA polymerases I and III. Proc Natl Acad Sci U S A 90:5524-8.
82. Langer, D., Hain, J., Thuriaux, P., Zillig, W. 1995. Transcription in archaea: similarity to that in eucarya. Proc Natl Acad Sci U S A 92:5768-5772.
83. Lebbink, J.H.G., De Vos, W.M. 1995. *Unpublished data.*
84. Leonard, P.M., S.H.J. Smits, S.E. Sedelnikova, A.B. Brinkman, W.M. de Vos, J. van der Oost, D.W. Rice, J.B. Rafferty 2001. Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. EMBO J 20:990-997
85. Leong, D., H. Boyer, and M. Betlach 1988. Transcription of genes involved in bacterio-opsin gene expression in mutants of a halophilic archaebacterium. J Bacteriol 170:4910-5.
86. Littlefield, O., Y. Korkhin, and P.B. Sigler 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. Proc Natl Acad Sci U S A 96:13668-13673.
87. Lopez-Garcia, P. 1999. DNA supercoiling and temperature adaptation: A clue to early diversification of life? J Mol Evol 49:439-52.
88. Lopez-Garcia, P., and P. Forterre 1997. DNA topology in hyperthermophilic archaea: reference states and their variation with growth phase, growth temperature, and temperature stresses. Mol Microbiol 23:1267-79.

89. Lopez-Garcia, P., S. Knapp, R. Ladenstein, and P. Forterre 1998. *In vitro* DNA binding of the archaeal protein Sso7d induces negative supercoiling at temperatures typical for thermophilic growth. Nucleic Acids Res 26:2322-8.
90. Luger, K., A. W. Mader, R.K. Richmond, D.F. Sargent, and T.J. Richmond 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389:251-60.
91. MacNeill, S.A. 2001. Understanding the enzymology of archaeal DNA replication: progress in form and function. Mol Microbiol 40:520-9.
92. Magill, C.P., S.P. Jackson, and S.D. Bell 2001. Identification of a conserved archaeal RNA polymerase subunit contacted by the basal transcription factor TFB. J Biol Chem 276:46693-6.
93. Murphy, S., B. Moorefield, and T. Pieler 1989. Common mechanisms of promoter recognition by RNA polymerases II and III. Trends Genet 5:122-6.
94. Musgrave, D., P. Forterre, and A. Slesarev 2000. Negative constrained DNA supercoiling in archaeal nucleosomes. Mol Microbiol 35:341-9.
95. Musgrave, D.R., K.M. Sandman, and J.N. Reeve 1991. DNA binding by the archaeal histone HMf results in positive supercoiling. Proc Natl Acad Sci U S A 88:10397-401.
96. Napoli, A., J. van der Oost, C.W. Sensen, R.L. Charlebois, M. Rossi, M. Ciaramella, 1999. An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. J Bacteriol 181:1474-1480.
97. Ng, W.V., S.P. Kennedy, G.G. Mahairas, B. Berquist, M. Pan, H.D. Shukla, S.R. Lasky, N.S. Baliga, V. Thorsson, J. Sbrogna, S. Swartzell, D. Weir, J. Hall, T.A. Dahl, R. Welti, Y.A. Goo, B. Leithauser, K. Keller, R. Cruz, M.J. Danson, D.W. Hough, D.G. Maddocks, P.E. Jablonski, M.P. Krebs, C.M. Angevine, H. Dale, T.A. Isenbarger, R.F. Peck, M. Pohlschroder, J.L. Spudich, K.W. Jung, M. Alam, T. Freitas, S. Hou, C.J. Daniels, P.P. Dennis, A.D. Omer, H. Ebhardt, T.M. Lowe, P. Liang, M. Riley, L. Hood, and S. DasSarma 2000. Genome sequence of *Halobacterium* species NRC-1. Proc Natl Acad Sci U S A 97:12176-81.
98. Nikolov, D.B., S.H. Hu, J. Lin, A. Gasch, A. Hoffmann, M. Horikoshi, N.H. Chua, R.G. Roeder, and S.K. Burley 1992. Crystal structure of TFIID TATA-box binding protein. Nature 360:40-6.
99. Opel, M.L., S.M. Arfin, and G.W. Hatfield 2001. The effects of DNA supercoiling on the expression of operons of the *ilv* regulon of *Escherichia coli* suggest a physiological rationale for divergently transcribed operons. Mol Microbiol 39:1109-15.
100. Palmer, J.R., and C.J. Daniels 1995. *In vivo* defenition of an archaeal promoter. J Bacteriol 177:1844-1849.
101. Pereira, S.L., R.A. Grayling, R. Lurz, and J.N. Reeve 1997. Archaeal nucleosomes. Proc Natl Acad Sci U S A 94:12633-7.
102. Pereira, S.L., and J.N. Reeve 1998. Histones and nucleosomes in Archaea and Eukarya: a comparative analysis. Extremophiles 2:141-8.
103. Perez-Martin, J., F. Rojo, and V. de Lorenzo 1994. Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. Microbiol Rev 58:268-90.
104. Pettijohn, D.E. 1996. The nucleoid, p. 158-166. In F.C. Neidhardt (ed.), *Escherichia coli* and Salmonella, cellular and molecular biology, vol. 1. ASM Press.
105. Qureshi, S.A., P. Baumann, T. Rowlands, B. Khoo, and S.P. Jackson 1995. Cloning and functional analysis of the TATA binding protein from *Sulfolobus shibatae*. Nucleic Acids Res 23:1775-81.

106. Qureshi, S.A., S.D. Bell, and S.P. Jackson 1997. Factor requirements for transcrip tion in the archaeon *Sulfolobus shibatae*. EMBO J 16:2927-2936.
107. Qureshi, S.A., and S.P. Jackson 1998. Sequence -specific DNA binding by the *S.shibatae* TFIIB homolog, TFB, and its effect on promoter strength. Mol Cell 1:389 -400.
108. Raumann, B.E., M.A. Rould, C.O. Pabo, and R.T. Sauer 1994. DNA recognition by beta -sheets in the Arc repressor-operator crystal structure. Nature 367:754 -7.
109. Reeve, J.N.,K. Sandman, and C.J. Daniels 1997. Archaeal histones, nucleosomes, and transcription initiation. Cell 89:999-1002.
110. Reiter, W.D.,U. Hudepohl, and W. Zillig 1990. Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start -site selection *in vitro*. Proc Natl Acad Sci U S A 87:9509-13.
111. Robinson, K.A., F.T. Robb, and H.J. Schreier 1994. Isolation of maltose -regulated genes from the hyperthermophilic archaeum, *Pyrococcus furiosus*, by subtractive hybridization. Gene 148:137 -41.
112. Robinson, K.A., and H.J. Schreier 1994. Isolation, sequence and characterization of the maltose-regulated *mlrA* gene from the hyperthermophilic archaeum *Pyrococcus furiosus*. Gene 151:173-6.
113. Roder, R., Pfeifer, F. 1996. Influence of salt on the transcription of the gas -vesicle genes of *Haloferax mediterranei* and identification of the endogenous transcriptional activator gene. Microbiology 142:1715 -1723.
114. Roeder, R.G. 1996. The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci 21:327-335.
115. Sandman, K., J.A. Krzycki, B. Dobrinski, R. Lurz, and J.N. Reeve 1990. HMf, a DNA -binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. Proc Natl Acad Sci U S A 87:5788-91.
116. Schiering, N., X. Tao, H. Zeng, J.R. Murphy, G.A. Petsko, and D. Ringe 1995. Structures of the apo - and the metal ion-activated forms of the diphtheria tox repressor from *Corynebacterium diphtheriae*. Proc Natl Acad Sci U S A 92:9843-50.
117. Schut, G.J., J. Zhou, and M.W. Adams 2001. DNA mic roarray analysis of the hyperthermophilic archaeon *Pyrococcus furiosus*: evidence for a new type of sulfur-reducing enzyme complex. J Bacteriol 183:7027-36.
118. Shand, R.F., and M.C. Betlach 1991. Expression of the *bop* gene cluster of *Halobacterium halobium* is induced by low oxygen tension and by light. J Bacteriol 173:4692 -9.
119. 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. Heik amp-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 U S A 98:7835-40.
120. Smale, S.T., A. Jain, J. Kaufmann, K.H. Emami, K. Lo, and I.P. Garraway 1998. The initiator element: a paradigm for core promoter heterogeneity withi n metazoan protein-endoding genes, p. 21-31, Mechanisms of transcription, vol. LXIII. Cold Spring Harbor Laboratory Press.

121. Soares, D., I. Dahlke, W.T. Li, K. Sandman, C. Hethke, M. Thomm, and J.N. Reeve 1998. Archaeal histone stability, DNA binding, and transcription inhibition above 90 degrees C. Extremophiles 2:75-81.
122. Soppa, J. 1999. Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box. Mol Microbiol 31:1589-1592.
123. Starich, M.R., K. Sandman, J.N. Reeve, and M.F. Summers 1996. NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. J Mol Biol 255:187-203.
124. Stetter, K.O. 1999. Extremophiles and their adaptation to hot environments. FEBS Lett 452:22-5.
125. Struhl, K. 1999. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell 98:1-4.
126. Taggart, A.K., and B.F. Pugh 1996. Dimerization of TFIID when not bound to DNA. Science 272:1331-3.
127. Thompson, D.K., Daniels, C.J. 1998. Heat shock inducibility of an archaeal TATA-like promoter is controlled by adjacent sequence elements. Mol Microbiol 27:541-551.
128. Thompson, D.K., J.R. Palmer, and C.J. Daniels 1999. Expression and heat-responsive regulation of a TFIIB homologue from the archaeon *Haloferax volcanii*. Mol Microbiol 33:1081-92.
129. Tyree, C.M., C.P. George, L.M. Lira-DeVito, S.L. Wampler, M.E. Dahmus, L. Zawel, and J.T. Kadonaga 1993. Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. Genes Dev 7:1254-65.
130. van der Oost, J., and S.D. Bell *Unpublished data.*
131. Verhees, C.H. 2002. Molecular characterization of glycolysis in *Pyrococcus furiosus*, Chapter 9: Promoter architecture of genes encoding glycolysis enzymes in *Pyrococcus furiosus*. Thesis.
132. Voorhorst, W.G., Y. Gueguen, A.C. Geerling, G. Schut, I. Dahlke, M. Thomm, J. van der Oost, and W.M. de Vos 1999. Transcriptional regulation in the hyperthermophilic archaeon *Pyrococcus furiosus*: coordinated expression of divergently oriented genes in response to beta-linked glucose polymers. J Bacteriol 181:3777-83.
133. Wang, J.C., and A.S. Lynch 1993. Transcription and DNA supercoiling. Curr Opin Genet Dev 3:764-8.
134. Washio, T., J. Sasayama, and M. Tomita 1998. Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. Nucleic Acids Res 26:5456-63.
135. Weis, L., and D. Reinberg 1997. Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. Mol Cell Biol 17:2973-84.
136. Werner, F., J.J. Eloranta, and R.O. Weinzierl 2000. Archaeal RNA polymerase subunits F and P are bona fide homologs of eukaryotic RPB4 and RPB12. Nucleic Acids Res 28:4299-305.
137. Wettach, J., H.P. Gohl, H. Tschochner, M. Thomm 1995. Functional interaction of yeast and human TATA-binding proteins with an archaeal RNA polymerase and promoter. Proc Natl Acad Sci U S A 92:472-476.
138. Woese, C.R., and G.E. Fox 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 74:5088-90.

139. Woese, C.R., O. Kandler, and M.L. Wheelis 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87:4576 -9.
140. Yang, C.F., J.M. Kim, E. Molinari, and S. DasSarma 19 96. Genetic and topological analyses of the *bop* promoter of *Halobacterium halobium*: stimulation by DNA supercoiling and non -B-DNA structure. J Bacteriol 178:840 -5.
141. Zhang, G., E.A. Campbell, L. Minakhin, C. Richter, K. Severinov, and S.A. Darst 1999. C rystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. Cell 98:811 -24.
142. Zhu, W., Q. Zeng, C.M. Colangelo, M. Lewis, M.F. Summers, and R.A. Scott 1996. The N -terminal domain of TFIIB from *Pyrococcus furiosus* forms a zinc ribbon. Nat Struct Biol 3:122 -4.
143. Zhulin, I.B., B.L. Taylor, and R. Dixon 1997. PAS domain S -boxes in Archaea, Bacteria and sensors for oxygen and redox. Trends Biochem Sci 22:331 -3.
144. Zillig, W., P. Palm, H.P. Klenk, D. Langer, U. Hüdepohl, J. Hain, M. Län zendörfer, I. Holz 1993. Transcription in archaea. In: Kates (ed.), The biochemistry of archaea. Elsevier Science Publishers B.V.
145. Zillig, W., K.O. Stetter, and D. Janekovic 1979. DNA -dependent RNA polymerase from the archaebacterium *Sulfolobus acidocaldarius*. Eur J Biochem 96:597 -604.
146. Zillig, W., K.O. Stetter, S. Wunderl, W. Schulz, H. Priess, I. Scholz 1980. The *Sulfolobus "Caldariella"* goup: taxonomy on de basis of the structure of DNA -dependent RNA polymerases. Arch Microbiol 125:259 -269.

# Chapter 2

## The archaeal TFIIEα homologue facilitates transcription initiation by enhancing TATA-box recognition

Stephen D. Bell
Arie B. Brinkman
John van der Oost
Stephen P. Jackson

**Transcription from many archaeal promoters can be reconstituted *in vitro* using recombinant TATA-box binding protein (TBP) and transcription factor B (TFB) - homologues of eukaryal TBP and TFIIB- together with purified RNA polymerase (RNAP). However, all archaeal genomes sequenced to date reveal the presence of TFE, a homologue of the α-subunit of the eukaryal general transcription factor, TFIIE. We show that, while TFE is not absolutely required for transcription in the reconstituted in vitro system, it nonetheless plays a stimulatory role on some promoters and under certain conditions. Mutagenesis of the TATA box or reduction of TBP concentration in transcription reactions sensitizes a promoter to TFE addition. Conversely, saturating reactions with TBP de-sensitizes promoters to TFE. These results suggest that TFE facilitates or stabilizes interactions between TBP and the TATA box.**

*EMBO Reports 2001 2(2): 133-138*

## INTRODUCTION

The archaeal basal transcription machinery is fundamentally related to the eukaryal RNA polymerase (RNAP) II machinery (5, 20, 22). A range of in *vitro* transcription assays have demonstrated that a variety of archaeal promoters of varying strengths can be accurately transcribed in reactions containing archaeal TATA-box binding protein (TBP), transcription factor B (TFB) and RNAP (3, 7, 9, 21). However, it has become apparent that all archaeal genomes sequenced to date encode a homologue of a third eukaryal general transcription factor, TFIIE (1, 4, 14). The archaeal TFIIE homologue, TFE, is related to the N-terminal 20 kDa of the α-subunit of TFIIE (Fig. 1A). Extensive analyses of eukaryal TFIIE function have been performed *in vivo* and *in vitro* (13, 19). In particular, deletion analyses have revealed that the C-terminal portion of TFIIEα is dispensable for viability in yeast but that the N-terminal portion is essential *in vivo* (13). Remarkably, the smallest truncated version of TFIIEα, which nonetheless permits yeast growth, corresponds closely to full-length archaeal TFE (Fig. 1A). In addition, similar deletion analyses that examined TFIIE function in reconstituted *in vitro* transcription assays revealed an analogous requirement for the N-terminal portion of human TFIIEα for basal and activated transcription (19). Thus, it appears that archaeal TFE corresponds to the minimal essential region of eukaryal TFIIEα (see Fig. 1A). Although the nature of the essential function remains undetermined, it has been demonstrated that this region of TFIIEα can interact with TBP and RNAP II (17, 18, 24). In archaea, the TFE open reading frame (ORF) contains an N-terminal, weakly conserved, helix-turn-helix motif within a leucine-rich region and a C-terminal Zn-ribbon (Fig. 1B). Intriguingly, while some archaeal TFEs possess putative Zn-ribbons with four cysteine residues coordinating the metal ion, those of *Sulfolobus solfataricus* and *Archaeoglobus fulgidus* have the second cysteine substituted by aspartic acid or methionine, respectively. While these maintain the potential to coordinate a metal ion via polar side chains, the TFEs of Pyrococci and *Methanobacterium thermoautotrophicum* possess proline or glycine in the analogous position (Fig. 1A). Thus, it is possible that not all archaeal TFEs possess a coordinated metal ion.

The reconstituted archaeal transcription assays performed to date demonstrate accurate transcription from a range of promoters in the presence of TBP, TFB and RNAP. Therefore, this indicates that TFE is not required for transcription from these promoters under the conditions employed. However, it is possible that TFE either plays a general stimulatory role in transcription or is required for a subset of promoters or under certain conditions. To determine whether TFE has a general role in archaeal transcription, we have

purified recombinant *S.solfataricus* P2 TFE and assayed its activity in a reconstituted *Sulfolobus in vitro* transcription system.

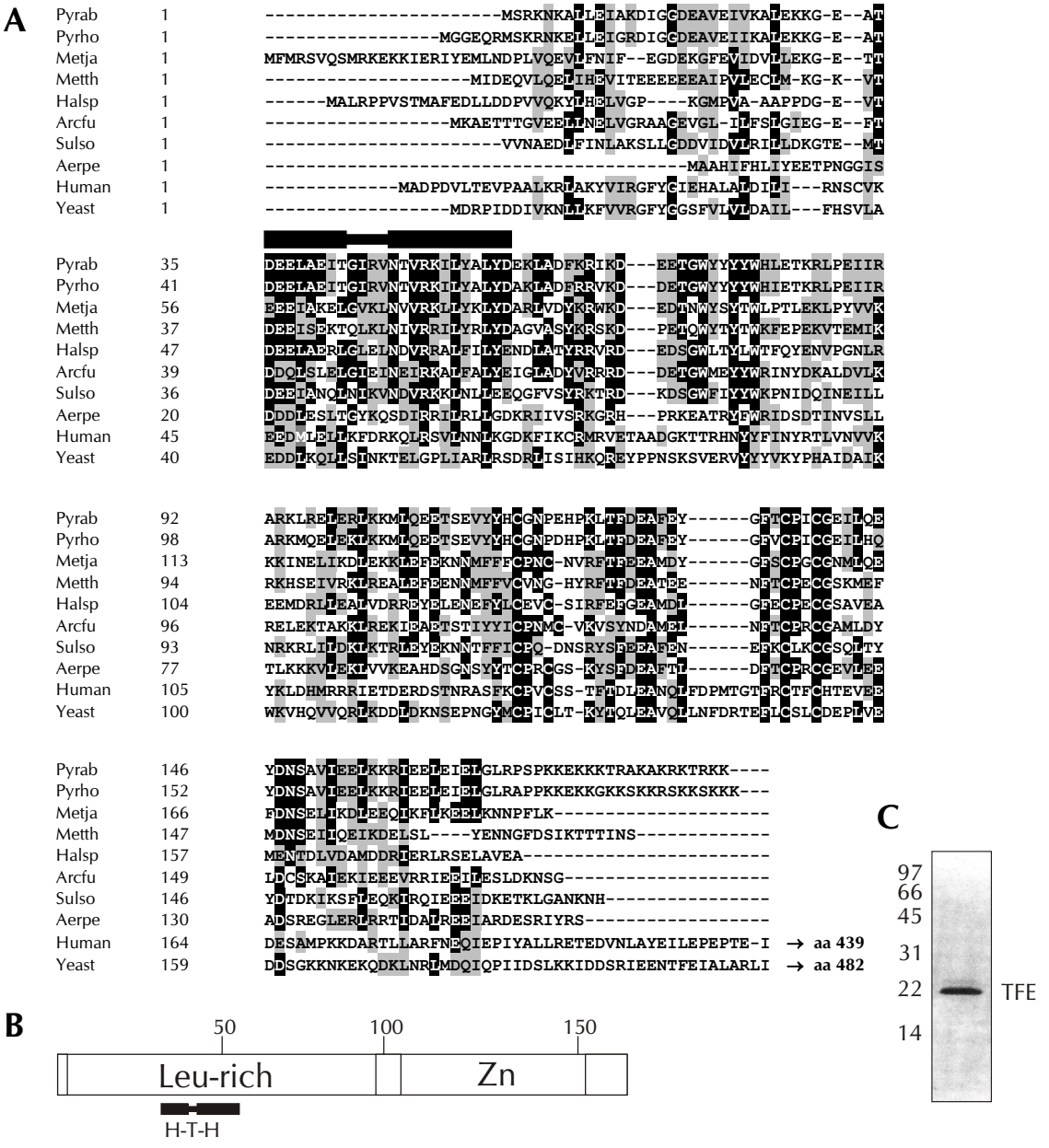



**Figure 1.** Archaea possess a sequence homologue of the α-subunit of TFIIE. (**A**) An alignment of various archaeal TFE and eukaryal TFIIEα subunits is shown. Pyrab, *Pyrococcus abyssi*, C75055; Pyrho, *Pyrococcus horikoshii*, B71106; Metja, *Methanococcus jannaschii*, Q58187; Metth, *Methanobacterium thermoautotrophicum*, A69090; Halsp, *Halobacterium sp*. NRC-1, AAG19231; Arcfu, *Archaeoglobus fulgidus*, E69344; Sulso, *Sulfolobus solfataricus*, http://niji.imb.nrc.ca/sulfolobus; Aerpe, *Aeropyrum pernix*, F72503; human, P29083; yeast, P36100. Identical residues are boxed in black and homologous residues are shaded in gray. A helix-turn-helix structure is indicated by black horizontal bars. (**B**) Schematic representation of motifs found in archaeal TFE. (**C**) Coomassie blue-stained SDS-polyacrylamide gel containing 3 μg of purified recombinant TFE.

## *RESULTS*

### Archaeal TFE interacts with RNAP and TBP

The ORF of *S.solfataricus* P2 TFE was identified in the *S.solfataricus* P2 genome sequence and amplified by PCR, followed by cloning and expression *in Escherichia coli* as a C-terminally $His_6$-tagged fusion protein. This protein was purified (Fig. 1C) and then used in protein-protein interaction assays to determine whether it could interact with components of the archaeal basal transcription machinery. First, 100 ng of TFE were mixed with 10 μg of *S.solfataricus* whole-cell extract, prior to incubation with Ni-NTA-agarose. Beads were washed extensively and bound protein eluted by boiling in SDS-PAGE loading buffer. Following electrophoresis, eluted proteins were detected by Western blotting using antisera generated against TBP, TFB or the B-subunit of RNAP (21). The interaction assays were carried out in the presence of 50 μg/ml ethidium bromide to exclude the possibility that interactions were bridged by DNA rather than being direct (15). The results of these assays show that TFE can interact with TBP and RNAP, but not TFB in whole-cell extracts (Fig. 2A). The assays were repeated using purified RNAP and purified recombinant TBP in place of extract (Fig. 2B). These assays confirmed the results of the pull-downs from extract, indicating that TFE can interact directly with TBP and RNAP in the absence of DNA.

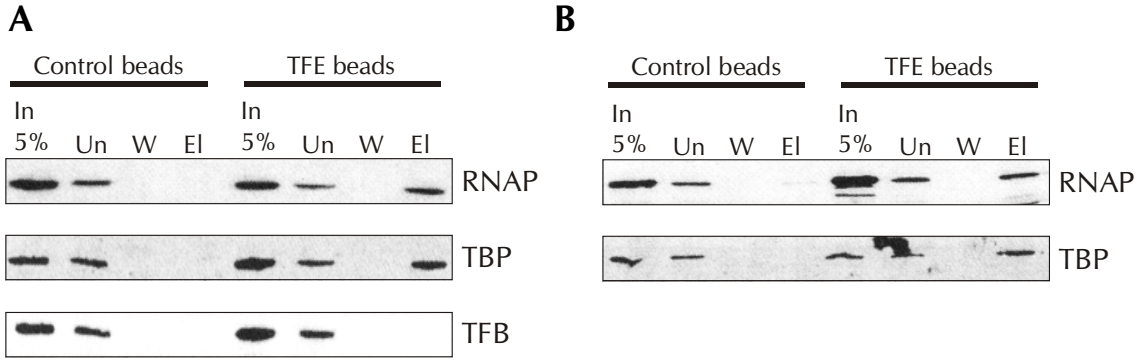



**Figure 2.** TFE interacts with RNAP and TBP. (**A**) Ni-NTA pull-down assays were performed with 10 μg of *S.solfataricus* extract, as detailed in Experimental Procedures, in the presence (TFE beads) or absence (control beads) of 100 ng recombinant TFE. The proteins pulled-down in the assay were detected by Western blotting with antisera raised against TBP, TFB or RNAP B-subunit (21). Lanes contain 5% of input (In), unbound material (Un), wash (W) or material eluted from the beads (El). (**B**) Assays performed as above but with purified TBP and RNAP in place of extract.

## TFE stimulates transcription on some promoters *in vitro*

To determine whether TFE plays a role in transcription, reconstituted transcription assays, using highly purified *Sulfolobus* RNAP and purified recombinant TBP and TFB, were performed on six different promoters, either with no added TFE or supplemented with 50 ng of TFE. As these assays contained RNAP purified from cells, Western blotting was performed using anti-TFE antisera to confirm that the RNAP preparation was essentially free of endogenous TFE (Fig. 3C). As seen in Figure 3A, under the reaction conditions employed, three promoters (*T6*, *lrs14* and *sdt*) are largely unaffected by TFE. However, the yield of transcript from three other promoters (*ArgC*, *5S* and *16S*) is stimulated two- to three-fold by the addition of TFE. Examination of the sequence of the core elements of these three TFE-responsive promoters reveals a common deviation from the TATA-box consensus in having a G at position 8 of the template strand of the TATA box, or the dyad-related C at position 1 of the element (Fig. 3B).

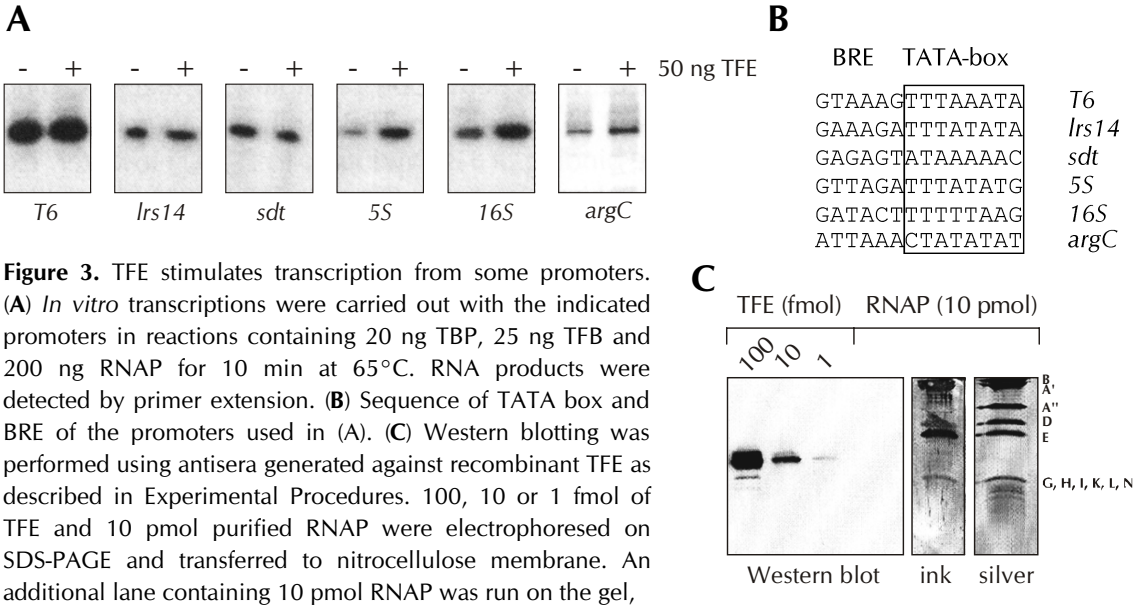



**Figure 3.** TFE stimulates transcription from some promoters. (**A**) *In vitro* transcriptions were carried out with the indicated promoters in reactions containing 20 ng TBP, 25 ng TFB and 200 ng RNAP for 10 min at 65°C. RNA products were detected by primer extension. (**B**) Sequence of TATA box and BRE of the promoters used in (A). (**C**) Western blotting was performed using antisera generated against recombinant TFE as described in Experimental Procedures. 100, 10 or 1 fmol of TFE and 10 pmol purified RNAP were electrophoresed on SDS-PAGE and transferred to nitrocellulose membrane. An additional lane containing 10 pmol RNAP was run on the gel, and excised prior to transfer. This was stained with silver and is shown in the right hand panel with the identity of protein bands indicated. The membrane containing immobilized TFE and RNAP was probed with the anti-TFE antisera, followed by goat anti-rabit IgG conjugated to horseradish peroxidase. Immunocomplexes were then detected by enhanced chemiluminesence and exposure to film (left hand panel). To confirm the RNAP had transferred to the membrane the membrane was stained with india ink (16) and photographed (note that subunits A" and D have negatively stained with this dye).

Previous work with human TFIIE has revealed that the α-subunit of TFIIE can stimulate the binding of TBP to the TATA element (24). We reasoned, therefore, that some non-consensus TATA-elements may confer sensitivity to TFE in our experiments. We note, however, that the *sdt* promoter, which appears to be unaffected by TFE addition, also

deviates from the TATA consensus. To test our hypothesis we made a series of guanine substitutions at positions 5, 6, 7 and 8 of the TATA box of the *T6* promoter, and tested the effect of TFE on transcription of these promoter derivatives (Fig. 4A and B). Somewhat surprisingly, substitution of guanine for adenine at positions 6 or 8 had only a small effect on transcription compared to wild-type *T6* promoter. Furthermore, these promoters were largely insensitive to the addition of TFE. In contrast, guanine substitution at position 7 and 5 resulted in a significant reduction of transcription relative to wild type. Importantly, these promoters were now stimulated by the addition of TFE to transcription reactions. This suggests that the relative affinity of TBP for the TATA box may impart sensitivity to TFE addition to a promoter.

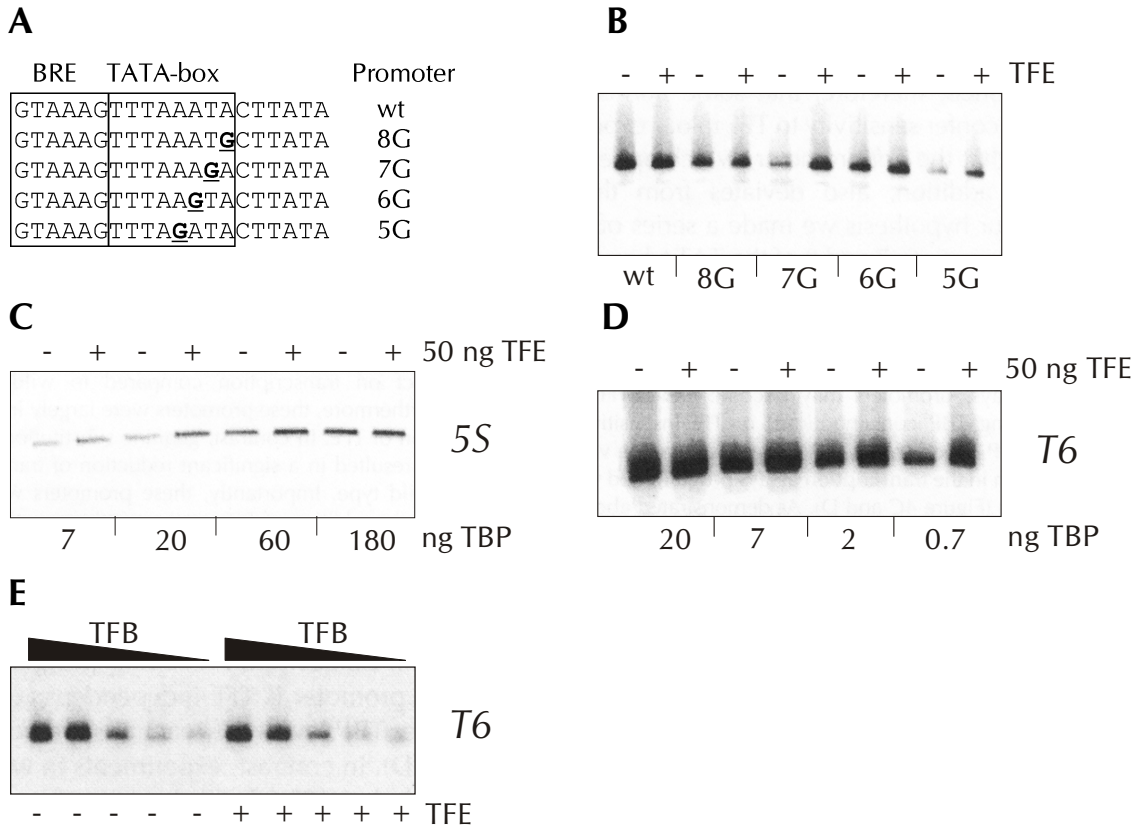



**Figure 4.** TFE stimulates transcription under sub-optimal TBP-TATA-box interactions. (**A**) The sequence of wild-type and mutant *T6* promoters is shown. The TATA box and BRE are boxed and the positions of sequence substitutions shown in bold and underlined. (**B**) The products of *in vitro* transcription of the promoters shown in (A) were detected by primer extension analysis and electrophoresed on an 8% denaturing polyacrylamide gel. (**C**) *In vitro* transcription reactions were performed on the *5S* promoter, either with no TFE or supplemented with 50 ng of TFE as indicated. The reactions contained 25 ng TFB, 200 ng of RNAP and varying amounts of TBP as indicated. (**D**) *In vitro* transcription reactions were performed on the *T6* promoter, either with no TFE or supplemented with 50 ng of TFE as indicated. Reaction conditions were as (C). (**E**) *In vitro* transcription assays performed on the *T6* promoter. Reactions contained 20 ng TBP, 200 ng RNAP and either 40, 20, 10, 5 or 2.5 ng TFB.

In light of the above, we reasoned that by altering TBP concentration in transcription assays, promoters may become either TFE-sensitive (under limiting TBP concentrations) or TFE-insensitive (under saturating TBP concentrations). To test this idea, we varied TBP concentration in the transcription assays programmed with *5S* or *T6* promoters (Fig. 4C and D). As demonstrated above, when the *5S* promoter is transcribed in reactions containing 20 ng of TBP, the addition of TFE stimulates transcription two-fold. However, as increasing TBP is added to a maximum of 180 ng the stimulatory effect is no longer detected (Fig. 4C). Similarly, we varied TBP levels in transcription reactions programmed by the *T6* promoter. The promoter is TFE-independent at 20 ng of TBP; however, at lower TBP levels the reactions are now stimulated by TFE (Fig. 4D). In contrast, experiments in which TFB concentration was varied and TBP levels kept constant at a saturating level showed no significant differences in the presence or absence of TFE (Fig. 4E). Together with the TATA-box substitution experiments in Figure 4, these data support the hypothesis that TFE can stimulate transcription under conditions where there are sub-optimal interactions between TBP and the TATA box.

## TFE stimulates early stages in transcription initiation

We next sought to determine whether TFE stimulates early stages in promoter recognition by performing DNaseI footprinting of TBP-TFB-DNA complex formation on the *T6* promoter at a range of TBP concentrations, in the presence or absence of TFE. As seen in Figure 5, TFE modestly enhances protection of the TATA-box region of the promoter; specifically, protection is observed at a two-fold lower TBP concentration in the presence of TFE. This level of stimulation is in agreement with the level of stimulation of transcription seen on the *T6* promoter under limiting TBP conditions (Fig. 4C).

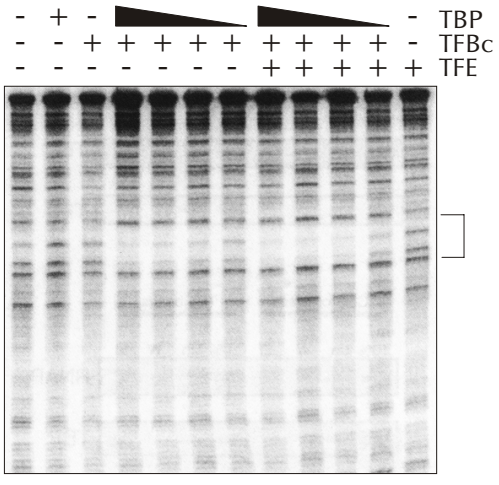



**Figure 5.** TFE stimulates TBP-TFB-DNA complex formation. DNaseI footprinting analysis was performed on the *T6* promoter in the presence or absence of 25 ng TFE and 20 ng TFBc (TFB core domain) as indicated. TBP was present at 20, 15, 10 or 5 ng per reaction (lanes 4-7 and 8-11).

## DISCUSSION

We find that TFE maximally exerts a two- to three-fold stimulatory effect on transcription, dependent on the promoter tested and reaction conditions. There appears to be a general correlation between detection of TFE-mediated stimulation and conditions where TBP-TATA-box interactions are sub-optimal. In agreement with this, we find that TFE interacts with TBP and that TFE appears to facilitate or stabilize recognition of the basal promoter elements by the basal machinery. How might this stimulation be mediated? One possibility is that TFE binds a DNA element in addition to TBP itself, and thereby facilitates cooperative binding of TBP and itself to DNA. However, extensive cross-linking analyses have been performed with the eukaryal basal transcription machinery, and there is no evidence for direct interactions between TFIIE and promoter DNA in the vicinity of the TATA box (10). Consistent with this, we observe that there is no detectable change in the TBP/TFB footprint observed in the presence or absence of TFE (Fig. 5). A second possibility is that TFE induces a conformational change in TBP, which facilitates DNA binding. This too seems unlikely, as the crystal structures of TBP in solution and bound to DNA have been determined and there is very little structural difference between the two forms of the protein (8, 12). A related possibility is suggested by the observation that TBP from eukarya and archaea exists as a dimer in solution. It has been proposed that dimerization of TBP prevents TBP-DNA interactions. It is possible that the function of TFE may prevent dimerization of TBP and so favor TATA-box recognition, indeed such behavior has recently been described for eukaryal TFIIA (6).

The effect that we observe with TFE is relatively weak: while withholding TBP or TFB from transcription assays essentially abolishes transcription on most promoters (21), the omission of TFE exerts a small quantitative effect *in vitro*. It is conceivable that a range of archaeal promoters exist that are more TFE dependent; based on our findings, we would propose that such promoters would contain TATA boxes that correspond poorly to the consensus sequence. In addition it is possible that *in vivo*, under physiological TBP concentrations, TFE plays a more significant effect. Additionally, in archaeal cells, DNA is compacted by interaction with a range of small basic proteins, including direct homologues of eukaryal histones in some species (22). Some of these proteins have been observed to repress transcription *in vitro* (23). It is possible that some of this repression results from competition for access to the TATA box between TBP and these non-specific DNA-binding proteins. If so, TFE could stimulate transcription *in vivo* by facilitating interaction

between TBP and the TATA box. Experiments to reconstitute archaeal chromatin *in vitro* are underway to address this and other related issues.

## ACKNOWLEDGEMENTS

We thank Jessica Downs for valuable comments on this manuscript. This work was supported by the Wellcome Trust

## EXPERIMENTAL PROCEDURES

### Cloning and purification of *S.solfataricus* P2 TFE

The ORF of TFE was identified by BLAST search of the publicly available *S.solfataricus* genome sequence (http://niji.imb.nrc.ca/sulfolobus) using the sequence of *A.fulgidus* TFE as query (11). The ORF was amplified by PCR using primers TFE5 (5'-GGGGAT-CCCATATGGTTAACGCAGAGGATCGTTT-3') and TFE3 (5'-GCCCTGACTCGA-GATGATTTTTATTAGCTCCAAG-3'). The PCR product was digested with *Nde*I and *Xho*I, and ligated to *Nde*I-*Xho*I digested pET30a (Novagen). The resultant expression plasmid pET-ssTFE was transformed into BLR RIL and expression of $His_6$-tagged TFE was induced during logarithmic growth of cells by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to 1 mM for 3 h. Cells were pelleted and resuspended in N300 (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM β-mercaptoethanol), lysed by sonication and clarified by centrifugation. The supernatant was heated to 70°C for 30 min. and re-centrifuged. The resultant supernatant had imidazole added to 20 mM and was then applied to a column containing Ni-NTA-agarose matrix (Qiagen). The matrix was washed with 10 column volumes of N300 + 45 mM imidazole, followed by elution by N300 + 500 mM imidazole. One-milliliter fractions were collected and the presence of TFE determined by SDS-PAGE and Coomassie blue staining. Positive fractions were pooled and dialyzed extensively against N300 + 10% glycerol + 10 μM $ZnSO_4$. Polyclonal antisera were generated in a rabbit against TFE (three consecutive injections of 100 μg TFE at three-week intervals). Serum was collected and tested for the presence of anti-TFE antibodies using standard methodologies (16).

## Purification of TBP, TFB and RNAP

These proteins were purified as described previously (3, 21).

## Transcription assays, EMSA and DNaseI footprinting

These were performed as described previously (2) using TBP, TFB, TFBc and RNAP purified as described (3).

## Protein-protein interaction studies

These were performed in a volume of 500 µl of BB (90 mM KCl, 20 mM Tris pH 8.0, 10 % glycerol, 10 mM β-mercaptoethanol, 50 µg/ml ethidium bromide, 0.1% Triton X-100, 20 mM imidazole, 10 µg/ml bovine serum albumin (BSA)). The reactions contained 100 ng purified TFE and either 10 µg *S.solfataricus* whole-cell extract, 1 µg of purified TBP or RNAP, as indicated in Figure 2. Reactions were shaken at room temperature together with 20 µl of a 50% slurry of Ni-NTA-agarose in BB. Beads were collected by centrifugation and washed 5 times with 1 ml of BB prior to resuspension in 40 µl 1x SDS-PAGE loading buffer and boiling. Eluted protein was subjected to SDS-PAGE and detected by Western blotting as described previously.

## REFERENCES


1. Aravind, L., E.V. Koonin 1999. DNA-binding proteins and evolution of transcription regulation in the archaea. Nucleic Acids Res 27:4658-70.
2. Bell, S.D., S.S. Cairns, R.L. Robson, S.P. Jackson 1999. Transcriptional regulation of an archaeal operon *in vivo* and *in vitro*. Mol Cell 4:971-82.
3. Bell, S.D., and S.P. Jackson 2000. The role of transcription factor B in transcription initiation and promoter clearance in the archaeon *Sulfolobus acidocaldarius*. J Biol Chem 275:12934-40.
4. Bell, S.D., and S.P. Jackson 1998. Transcription in Archaea, p. 41-51, Mechanisms of transcription, vol. LXIII. Cold Spring Harbor Laboratory Press.
5. Bell, S.D., S.P. Jackson 1998. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends in Microbiology 6:222-8.
6. Coleman, R.A., A.K. Taggart, S. Burma, J.J. Chicca 2nd, and B.F. Pugh 1999. TFIIA regulates TBP and TFIID dimers. Mol Cell 4:451-7.
7. Darcy, T.J., W. Hausner, D.E. Awery, A.M. Edwards, M. Thomm, and J.N. Reeve 1999. *Methanobacterium thermoautotrophicum* RNA polymerase and transcription *in vitro*. J Bacteriol 181:4424-9.
8. DeDecker, B.S., R. O'Brien, P.J. Fleming, J.H. Geiger, S.P. Jackson, and P.B. Sigler 1996. The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. J Mol Biol 264:1072-84.
9. Hethke, C., A.C. Geerling, W. Hausner, W.M. de Vos, and M. Thomm 1996. A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. Nucleic Acids Res 24:2369-2376.
10. Kim, T.K., R.H. Ebright, and D. Reinberg 2000. Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science 288:1418-22.
11. Klenk, H.P., R.A. Clayton, J.F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, D.L. Richardson, A.R. Kerlavage, D.E. Graham, N.C. Kyrpides, R.D. Fleischmann, J. Quackenbush, N.H. Lee, G.G. Sutton, S. Gill, E.F. Kirkness, B.A. Dougherty, K. McKenney, M.D. Adams, B. Loftus, J.C. Venter, *et al.* 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature 390:364-70.
12. Kosa, P.F., Ghosh, G., B.S. DeDecker, P.B. Sigler 1997. The 2.1-Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. Proc Natl Acad Sci U S A 94:6042-47.
13. Kuldell, N.H., and S. Buratowski 1997. Genetic analysis of the large subunit of yeast transcription factor IIE reveals two regions with distinct functions. Mol Cell Biol 17:5288-98.
14. Kyrpides, N.C., C.A. Ouzounis 1999. Transcription in archaea. Proc Natl Acad Sci U S A 96:8545-50.
15. Lai, J.S., and W. Herr 1992. Ethidium bromide provides a simple tool for identifying genuine DNA-independent protein associations. Proc Natl Acad Sci U S A 89:6958-62.
16. Lane, D., and E. Harlow 1988. Antibodies, a laboratory manual. Cold Spring Harbor Laboratory Press, New York.
17. Maxon, M.E., J.A. Goodrich, and R. Tjian 1994. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev 8:515-24.
18. Maxon, M.E., and R. Tjian 1994. Transcriptional activity of transcription factor IIE is dependent on zinc binding. Proc Natl Acad Sci U S A 91:9529-33.

19. Ohkuma, Y., S. Hashimoto, C.K. Wang, M. Horikoshi, and R.G. Roeder 1995. Analysis of the role of TFIIE in basal transcription and TFIIH-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-alpha. Mol Cell Biol 15:4856-66.
20. Orphanides, G., T. Lagrange, and D. Reinberg 1996. The general transcription factors of RNA polymerase II. Genes Dev 10:2657-83.
21. Qureshi, S.A., S.D. Bell, and S.P. Jackson 1997. Factor requirements for transcription in the archaeon *Sulfolobus shibatae*. EMBO J 16:2927-36.
22. Reeve, J.N., K. Sandman, and C.J. Daniels 1997. Archaeal histones, nucleosomes, and transcription initiation. Cell 89:999-1002.
23. Soares, D., I. Dahlke, W.T. Li, K. Sandman, C. Hethke, M. Thomm, and J.N. Reeve 1998. Archaeal histone stability, DNA binding, and transcription inhibition above 90 degrees C. Extremophiles 2:75-81.
24. Yokomori, K., C.P. Verrijzer, and R. Tjian 1998. An interplay between TATA box-binding protein and transcription factors IIE and IIA modulates DNA binding and transcription. Proc Natl Acad Sci U S A 95:6722-7.

# Chapter 3

## The prokaryotic Lrp family of transcriptional regulators

Arie B. Brinkman
Thijs J.G. Ettema
Willem M. de Vos
John van der Oost

**The Lrp family of transcriptional regulators is generally involved in the control of amino acid metabolism in prokaryotes, both bacteria and archaea. Although the regulon of the archetype Lrp of *Escherichia coli* includes dozens of genes, it appears that the majority of Lrp-like proteins specifically controls the expression of a single gene or operon. The recently solved three-dimensional structure of *Pyrococcus furiosus* LrpA has revealed a classical helix-turn-helix DNA-binding domain, which is connected with a hinge to a regulatory domain that constitutes the effector binding site. The specific activity of Lrp-like proteins appears to be modulated allosterically by amino acid effectors, affecting oligomerization and DNA-binding characteristics. A sequence profile analysis has revealed that the Lrp regulatory domain, called RAM, is not only present in transcriptional regulators, but also fused to some metabolic enzymes and as isolated modules. Although the structure and function of the RAM-domain strongly resembles that of the ubiquitous ACT domain, it is discussed that their effector binding sites are different, most likely reflecting convergent evolution of these regulatory modules.**

## INTRODUCTION

Lrp (leucine-responsive regulatory protein) from *E.coli* is the prototype of the DNA-binding Lrp family of transcriptional regulators, termed after the 'leucine effect' that it mediates by affecting the expression of many different operons in the response to leucine in the growth medium. *E.coli* Lrp has been the focus of many studies because of its global role in transcriptional regulation of several dozens of operons involved in a variety of cellular processes that respond to the availability of amino acids in the medium (13, 50). Regulation by *E.coli* Lrp follows different patterns: Lrp may either activate or repress transcription, while leucine (or alanine) either stimulates or reduces this effect, or has no effect at all (13). In *E.coli*, Lrp often acts in concert with other (global) regulators like CAP (catabolite activator proteins), IHF (integration host factor), or the histone-like protein H-NS (37, 45, 54, 70). Furthermore, Lrp blocks DNA methylation of some *E.coli* promoters (66, 71). Although *E.coli* Lrp is by far the best-studied member of the Lrp family, a number of additional Lrp-like proteins have been studied in the past decade, some of which in considerable detail. In this review, we aim at a comparison of these Lrp-like proteins in order to describe general characteristics of their physiological function, as well as their mechanism of regulation. In addition, the recently solved three-dimensional structure of *P.furiosus* LrpA allows for addressing structure-function relations. Moreover, a new definition of Lrp-like proteins is provided, which may contribute to an improved annotation of genes encoding Lrp-like proteins.

## GLOBAL AND SPECIFIC FUNCTIONS

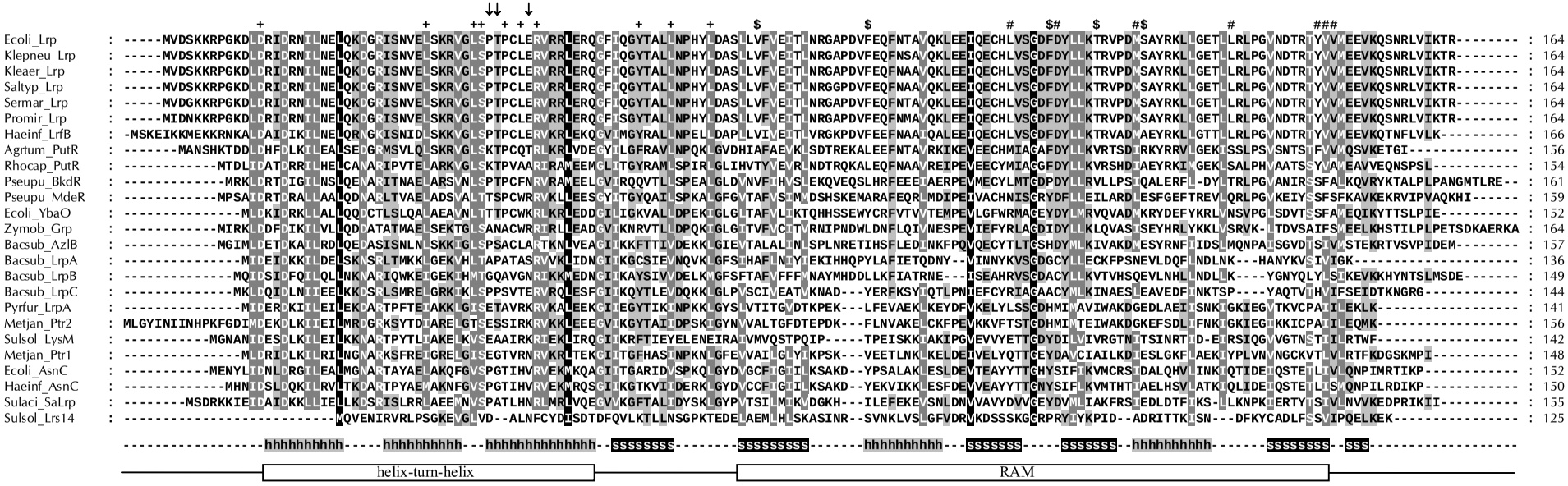
The recent availability of many complete genome sequences has made clear that Lrp-like proteins are widely spread among prokaryotes, bacteria as well as archaea. An analysis of all prokaryotic genomes sequenced to date reveals that obvious homologues of Lrp-like proteins are present in 17 out of 33 bacterial genera and all 14 archaeal genera. In contrast, no Lrp-like proteins have yet been found in eukarya. The amount of Lrp-paralogues per genome varies, but the highest number encountered so far is 20 in *Mesorhizobium loti*. Obviously, in organisms containing several paralogues some could have merely specific functions, whereas others have a more (moderate) global role. For example, *E.coli* contains three Lrp/AsnC paralogues, Lrp, AsnC, and an open reading frame (ORF) called *ybaO* (10). Lrp is a global regulator, whereas AsnC is specific for regulation of its own transcription

and that of *asnA*, encoding asparagine synthetase A (34). YbaO has not yet been characterized experimentally, but BLAST search analysis showed that it is most closely related to a protein called Grp (74), which was identified as a regulator for glutamate uptake in *Zymomonas mobilis* (Fig. 1).

Recently, Friedberg *et al.* showed that Lrp-like proteins with high identity to *E.coli* Lrp (>97%) are restricted to enteric bacteria, whereas other Lrp's have lower pair-wise sequence identity (29% on average) (24). This is also evident from Figure 1B, which shows that Lrp's from enteric bacteria are closely related, but do not cluster with any of the other characterized Lrp-like proteins from bacterial or archaeal origin. It has been proposed that Lrp's in enteric bacteria are highly adapted for their global function, and that only few amino acid substitutions are therefore tolerated, in agreement with the variety of functions and interactions that these proteins are involved in (25). Lrp-like proteins with lower sequence identity than 97% are thus less likely to be global regulators. As a case study to test this idea, the role *Haemophilus influenzae* LrfB was investigated. Although LrfB is one of the most closely *E.coli*-Lrp related proteins of non-enteric origin (75% identity), expression of only two proteins was obviously affected in an *H.influenzae lrfB*$^-$ mutant, suggesting a specific role for this protein (24). In contrast, similar experiments performed with *E.coli lrp*$^-$ mutants showed that Lrp affects expression of 30 to 75 proteins (21, 38). The authors propose the idea that only the enteric bacteria have a globally acting Lrp, and that Lrp-like proteins in other non-enteric bacteria may have rather specific, or moderate global functions. It is thought that the global function of Lrp's in enteric bacteria is to help these organisms adjust their metabolism to the nutritionally rich or poor conditions that they encounter inside the intestinal tract or outside the host, respectively (13, 49). However, since large fluctuations in substrate availability are expected to occur in other ecosystems as well, it is conceivable that non-enteric bacteria potentially utilize Lrp-like proteins as global regulators a well. Since these might have evolved independently from *E.coli* Lrp, such regulators might lack high sequence identity to *E.coli* Lrp.

Because the number of Lrp-paralogues of archaea is comparable to that of bacteria, it is likely that Lrp's in bacteria and archaea have comparable functions, i.e. mainly specific or moderate global. The role of archaeal Lrp-like proteins has thus far only been studied *in vitro*, and it is therefore not possible yet to determine whether these proteins have global or only specific functions (7, 12, 20, 48).

**A**




**B**

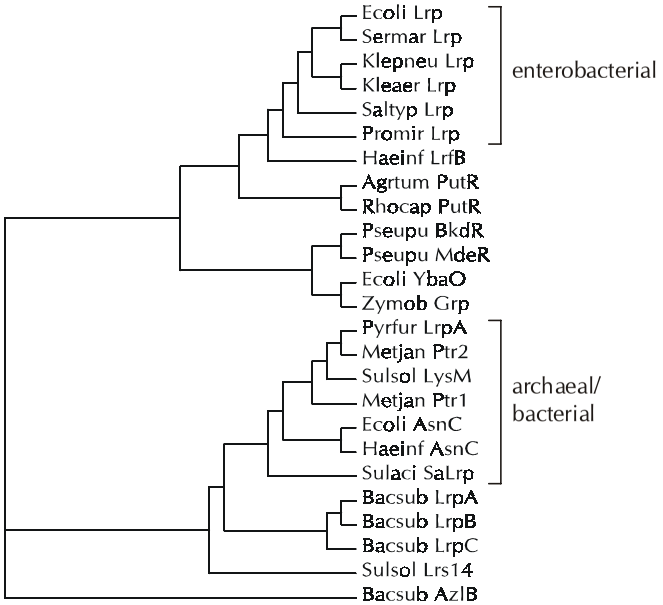



**Figure 1.** (**A**) Alignment of Lrp-like proteins. Secondary and domain structure of LrpA is indicated at the bottom; **hh**, α-helix; ss, β-sheet. Different classes of mutations in *E.coli* Lrp are indicated; +, DNA-binding; $, activation; #, leucine response. Vertical arrows refer to E32, T33, and K37 of *P.furiosus* LrpA. These residues correspond with the base-specific DNA-contacting residues R180, E181, and R185 of *E.coli* CAP in a superimposition with the LrpA and CAP HTHs (see text). (**B**) Dendrogram of Lrp-like proteins. Ecoli_Lrp, P19494; Klepneu_Lrp, P37424; Kleaer_Lrp, O87635; Saltyp_Lrp, P37403; Sermar_Lrp, 407921 Promir_Lrp, CAA71443.1; Haeinf_LrfB, NP_439738.1; Agrtum_PutR, Q44333; Rhocap_PutR, Q52710; Pseupu_BkdR, P42179; Pseupu_MdeR, 2217944; Ecoli_YbaO, P54986; Zymob_Grp, P74996 Bacsub_AzlB, O07920; Bacsub_LrpA, P96652; Bacsub_LrpB, P96653; Bacsub_LrpC, P96582 Pyrfur_LrpA, P42180; Metjan_Ptr2, Q58133; Sulsol_LysM, Sso0157; Metjan_Ptr1, Q57615; Ecoli_AsnC, P03809; Haeinf_AsnC, P44337; Sulaci_SaLrp, T45215; Sulsol_Lrs14, AAD19358.1.

| organism | Lrp-like protein | global / specific | ligand(s) | identified target genes | regulatory effect | function | % identity to *E.coli* Lrp | references |
|---|---|---|---|---|---|---|---|---|
| *Escherichia coli* | Lrp | global | leucine/ alanine | 35 to 75 different genes | activation/ repression | amino acid biosynthesis, amino acid degradation, transport, pili formation | 100 | (13, 21, 38, 50), and references therein |
| *Salmonella typhimurium* | Lrp | global[1] | leucine[1] | *ilvIH*<br>*pef*<br>*spv* | activation/ repression[1] | branched-chain amino acid biosynthesis<br>pili formation<br>plasmid virulence | 98 | (25, 44, 51, 69) |
| *Proteus mirabilis* | Lrp | global[1] | leucine[1] | *flhDC* | activation/ repression[1] | flagellation and swarming | 97 | (28, 35) |
| *Escherichia coli* | AsnC | specific[2] | asparagine | *asnA, asnC* | activation/ repression | asparagine biosynthesis | 25 | (19, 33, 34) |
| *Haemophilus influenzae* | LrfB | specific | ? | *asnA, hyp. protein*[4] | activation/ repression[4] | asparagine biosynthesis | 75 | (24) |
| *Pseudomonas putida* | BkdR | specific | L-valine, L-isoleucine, L-leucine, D-leucine | *bkdR, bkdAB* | activation/ repression | catabolism of branched-chain amino acids | 38 | (39, 40, 42, 43) |
| *Pseudomonas putida* | MdeR | specific[2] | ? | *mdeAB* | activation | catabolism of methionine | 34 | (29) |
| *Agrobacterium tumefaciens* | PutR | specific[2] | proline | *putA, putR* | activation/ repression | catabolism of proline | 48 | (15, 30) |
| *Rhodobacter capsulatus* | PutR | specific[2] | proline | *putA, putR* | activation/ repression | catabolism of proline | 45 | (31) |
| *Bacillus subtilis* | LrpA / LrpB | ? | ? | *glyA* | ? | serine-glycine interconversion, KinB-dependent sporulation | 27 / 28 | (18) |
| *Bacillus subtilis* | LrpC | ? | ? | *lrpC* | activation | sporulation, amino acid metabolism, positive autoregulation | 32 | (8, 9) |
| *Bacillus subtilis* | AzlB | specific[2] | ? | *azlBCDEF* | repression | branched-chain amino acid transport | 33 | (5) |
| *Sulfolobus solfataricus* | LysM | specific[2] | lysine | *lysWXJK* | activation | biosynthesis of lysine | 21 | (11) |
| *Sulfolobus solfataricus* | Lrs14 | ? | ? | *lrs14* | repression | only negative autoregulation shown | NS[3] | (7, 48) |
| *Sulfolobus acidocaldarius* | SaLrp | ? | ? | ? | ? | ? | 29 | (20) |
| *Pyrococcus furiosus* | LrpA | ? | ? | *lrpA* | repression | only negative autoregulation shown | 28 | (12) |
| *Methanococcus jannaschi* | Ptr1/Ptr2 | ? | ? | *ptr1 / ptr2* | ? | ? | 28 / 24 | (53) |
| *Zymomonas mobilis* | Grp | specific[2] | ? | *gluEMP, grp* | repression | glutamate uptake | 35 | (55, 56) |

**Table I.** Characteristics of studied Lrp/AsnC proteins

[1] It is assumed that these Lrp homologues have the same function as Lrp in *E.coli*.

[2] Clustering with its target gene strongly suggest a specific role, although there is no direct (genetic) evidence excluding a global regulatory function.

[3] NS, no similarity (see text)

[4] As determined by changes in spot-intensity using two-dimensional gel-electrophoresis.

## GENERAL FEATURES

General characteristics of Lrp-like proteins that have been characterized in more detail have been summarized in Table I. In bacteria, identified target genes for Lrp-like proteins are involved in amino acid metabolism, and their ligands, when identified, turned out to be amino acids (Table I). Given the overall sequence identity between bacterial and archaeal Lrp-like proteins it is to be expected that archaeal Lrp-like proteins have similar functions in regulation of amino acid metabolism. Unfortunately, for most of the studied archaeal Lrp's no target genes have been identified, and due to the genetic inaccessibility of most archaea, no archaeal *lrp*$^-$ mutants are yet available. Hence, studies on archaeal Lrp-like proteins have been based on comparative genomics rather then on genetic manipulation, and so far only ligand-independent negative autoregulation could be demonstrated in *in vitro* experiments (7, 12). The first example of an archaeal Lrp-like protein for which a ligand and target promoter have been established concerns LysM from the hyperthermophilic archaeon *Sulfolobus solfataricus* (see Table I). The gene encoding LysM is clustered with genes that are involved in the biosynthesis of lysine, and transcription from one of the cluster's promoters is strongly regulated by the presence of lysine in the medium. *In vitro* and *in vivo* data demonstrated that LysM occupies its binding site upstream of the promoter preferably in the absence of lysine, strongly suggesting that LysM activates transcription from this promoter, and that lysine acts as a ligand (11). Together, these studies indicate that archaeal Lrp's, like bacterial Lrp's, are potential activators or repressors involved in the regulation of amino acid metabolism, and that amino acids serve as their ligands. Hence, it appears that Lrp-like proteins are functionally equivalent in the bacterial and archaeal domains, despite the fundamental differences in transcriptional machineries.

An intriguing question is how amino acids modulate the regulatory effect of Lrp-like proteins. Analysis of random *E.coli* Lrp mutants has shown that the C-terminal part of the protein is involved in leucine response (58), and a number of effects of amino acid ligands on Lrp-like proteins have been observed. (I) They affect DNA binding affinity (11, 22, 30, 39, 59). While the *in vitro* effect on DNA-binding may be small and effective only within a narrow Lrp concentration range, it could have a major effect *in vivo*. (II) They alter Lrp-induced DNA bending (13, 30, 39). Either the bend angle or the bend center can be affected. Ligand-induced changes in the DNaseI footprinting pattern may be related to this phenomenon. (III) They cause a conformational change in the protein, as determined by the susceptibility to limited proteolysis. (40). (IV) They affect oligomerization, as was

shown recently for *E.coli* Lrp by a leucine-promoted dissociation of a hexadecameric to an octameric form (see also below) (14).

As noted above the effect of amino acid ligands on regulation by Lrp-like proteins is different among target promoters. This flexibility in amino acid response is not only restricted to *E.coli* Lrp, but rather appears to be a general feature of Lrp-like proteins. Usually, Lrp-like proteins negatively regulate their own transcription independently of their respective ligands, whereas activation of their target genes often requires the ligands (30, 31, 34, 41, 42). In some cases the location of the Lrp binding site(s) with respect to its target promoter is an indication for its regulatory effect. Characteristic examples are binding sites within intergenic regions containing two divergent promoters, one driving transcription of the gene encoding the Lrp-like proteins itself, and the other driving transcription of its target gene(s). Typically, the target gene is positively regulated in a ligand-dependent manner, whereas the *lrp* gene itself is negatively autoregulated, independent of the amino acid ligand. The two promoters share the same Lrp target sequence, but the effect is opposite for the two promoters. BkdR and PutR are examples of such a situation (Fig. 2A).

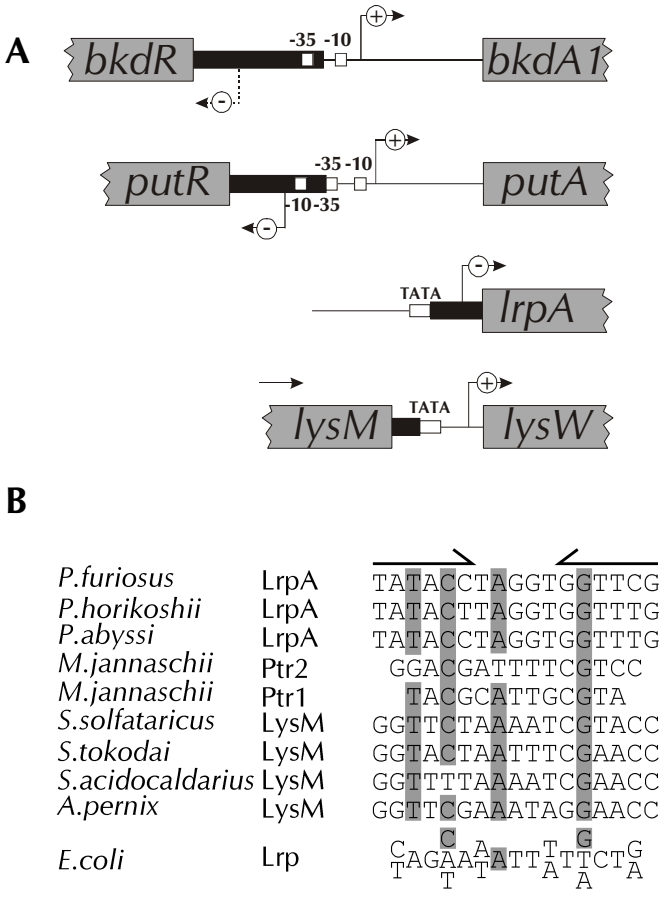



**Figure 2.** (**A**) Locations of binding sites for bacterial and archaeal Lrp-like proteins. Arrows indicate transcription start sites and the effect of the respective Lrp-like protein (+ or -). White rectangles indicate promoter elements, and solid black bars indicate binding sites for BkdR, PutR, LrpA, and LysM, as determined by footprinting experiments. The exact transcriptional start of *bkdR* is not known (dashed arrow). Transcription of *lysM* is driven from the more upstream *lysY* promoter. (**B**) Alignment of binding site sequences for Lrp-like proteins. The binding sites for *M.jannaschii* Ptr1, Ptr2, and *E.coli* Lrp were determined by SELEX (16, 53, 65); binding sites for *P.furiosus* LrpA and *S.solfataricus* LysM were determined using footprinting (11, 12); binding sites for their paralogues were determined by comparison of the respective promoter sequences. Horizontal arrows indicate (partial) inverted repeat elements.

Both proteins bind to sequences that overlap most of the -35 to -10 region of the repressed promoters (*bkd*R and *put*R), but not the activated promoters (*bkdAB* and *putA*) (30, 41). This suggests that repression here is a result of promoter-occupation, and as such preventing RNA polymerase binding. Although the organization of the *E.coli asnA* and *asnC* genes is most likely similar (34), no AsnC-DNA binding studies have been reported yet.

Variations on the theme of repression by promoter-occlusion are found in archaea, which have a eukaryal-like basal transcription machinery (6). In contrast to the bacterial RNA polymerase and sigma factor (RNAP holoenzyme), which bind directly to the -35 and -10 promoter elements, archaeal transcription initiation is preceded by the binding of TATA-binding protein (TBP) to the TATA element, followed by the binding of transcription factor B (TFB), which in turn recruits RNAP. LrpA from the hyperthermophilic archaeon *P.furiosus* binds downstream of the TATA element of the *lrpA* promoter, overlapping the transcriptional start. This blocks recruitment of RNAP, but not formation of the ternary DNA-TBP-TFB complex (12, 17).

In contrast to repression by Lrp-like proteins, the exact mechanism for activation by Lrp-like proteins is not completely understood. Although mutations affecting activation by *E.coli* Lrp map mainly in the central and C-terminal part of the protein (58), direct contacts between this region of Lrp and RNA polymerase have not yet been demonstrated. Since Lrp-like proteins have shown to be capable of changing DNA secondary structure (see below), their role in activation might as well be architectural. For instance, BkdR activates transcription from the *bkd* operon only in the presence of the inducing ligands L-valine, L-isoleucine, L-leucine and D-leucine (43). Besides a small effect on DNA binding affinity, a conformational change in BkdR and an altered bend angle of the BkdR-DNA complex is induced by L-valine (39, 40). It has therefore been suggested that activation depends on BkdR-induced promoter remodeling, thereby enhancing RNA polymerase binding and initiation of transcription. It is unclear yet whether such mechanisms are common to all Lrp-like proteins, but since they are independent of direct interactions with RNA polymerase, this may explain how Lrp-like proteins could achieve activation in both bacterial and archaeal domains, where basal transcriptional machinery's are fundamentally different (6). On the other hand, bacterial and archaeal Lrp's might have diverged sufficiently to evolve specific functions for activating either bacterial or archaeal transcription.

Sequences of naturally occurring binding sites for Lrp-like proteins often lack perfect inverted repeat elements. Compared to the binding sites of global regulators such as

catabolite activating protein (CAP) and fumarate and nitrate reduction regulator (FNR), the recognition sequences of which appear to be almost perfect inverted repeats (4), binding sites of Lrp-like proteins cannot be easily distinguished in target promoters. Development of *in vitro* evolved optimal binding sites for Lrp-like proteins using systematic evolution of ligands by exponential enrichment (SELEX) (65) provides a better way to screen for sequence preferences. Such experiments have shown that individual high-affinity binding sequences for Lrp's do indeed contain sequence elements with dyad symmetry, and that specific bases are preferred at some positions (16, 53) (Fig. 2B). Some wild-type Lrp-binding sites of hyperthermophilic archaea show a remarkable homology with SELEX-derived sequences. In particular, a palindromic C/G base couple is virtually conserved in these binding sites, while the wild-type binding sites for LysM appear to have a conserved palindromic GGTTC element.

It has been suggested that cooperativity in binding to multiple sub-optimal binding sites overcomes the lack of a single optimal binding site, and therefore allows for high affinity binding (16). This is in agreement with most of the data obtained from footprinting experiments with Lrp-like proteins, where stretches of hundred bp or more are typically protected against cleavage (7, 12, 30, 41, 52, 67, 72). In addition, Lrp-like proteins bend their target DNA sequences, as apparent from *in vitro* bending assays and the induction of DNaseI hypersensitive cleavage sites in footprinting experiments. Sequence-dependent deformability of DNA has been shown to contribute to the specificity and strength of a protein-DNA interaction (64). *In vitro* bending assays have revealed that *E.coli* Lrp, *Pseudomonas putida* BkdR and *Agrobacterium tumefaciens* PutR bend their target DNA with angles ranging from 52° to 135° (30, 39, 68), and possibly an even more drastic distortion of DNA structure occurs. For PutR this was confirmed experimentally using atomic force microscopy, which showed that the protein condenses more than 100 bp of its target DNA into a nucleoprotein complex (30), suggesting DNA wrapping around the protein.

## STRUCTURE OF LrpA

LrpA from *P.furiosus* is the first Lrp-like protein to date of which a three-dimensional structure has been solved (36). In the crystal structure LrpA forms an octamer consisting of four dimers (Fig. 3). The structure revealed that the N-terminal part of the protein consists of a helix-turn-helix (HTH) domain, a fold generally responsible for DNA binding. The HTH fold had previously been predicted for several Lrp-like proteins by computer-based

analysis of the primary amino acid sequence, and the presence of such a domain in the N-terminal part of the protein is in agreement with the observation that N-terminal mutations in several Lrp-like proteins affect DNA-binding (20, 53, 58). HTH domains are not only present in many regulatory proteins from prokaryotes, but also in eukaryal DNA-binding proteins, such as homeodomains and (heat shock) transcription factors. The structure of the HTH domain from LrpA closely resembles that of other prokaryotic regulators like CAP (46), and tryptophan repressor (TrpR) (60). In LrpA the HTH forms a distinct 'headpiece' of the protein, connected with a hinge to its C-terminal domain.

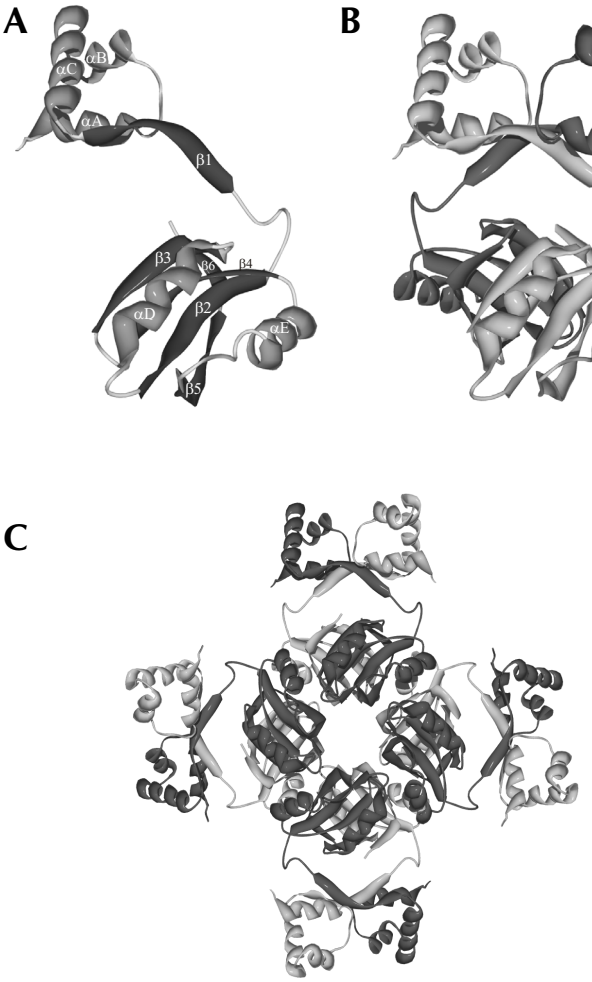



**Figure 3.** Three-dimensional structure of the *P.furiosus* LrpA monomer (**A**), dimer (**B**), and octamer (**C**), modified from Leonard et *al*. (36). LrpA consists of an N-terminal HTH domain (αA,αB,αC) and a C-terminal RAM domain (β2,αD,β3,β4,αE,β5), connected by a hinge (β1).

Although it is clear that the HTH-domain of Lrp-like proteins facilitates DNA binding, it is not known which of the amino acid residues in the recognition helix (αC, Fig, 3) determine the sequence specificity of DNA binding. To identify residues in LrpA involved in base-specific protein-DNA contacts, we applied a structure-based prediction method (73), in which the HTH structure is assigned to one of the described structural HTH families based on the spatial arrangement of the three helices of the motif. Subsequently, equivalent residues can be identified and predicted to have similar interactions with DNA. Obviously, this requires that the structural family contains at least one member whose structure in complex with DNA has been solved. Based on this method, LrpA can be assigned to the CAP-family of HTH

domains. Although the HTH of CAP is actually a winged-HTH, containing a two-stranded β-hairpin inserted between helix A and B, the spatial arrangement of all three helices is very similar to that of the three helices in the LrpA HTH: 30 Cα atoms of the 42-residue LrpA HTH motif can be superimposed on the CAP HTH with a root mean square deviation (RMSD) of 0.94 Å (Fig. 4). This value is the lowest RMSD value among members of the CAP-HTH structural family: the HTHs of other members like BirA and LexA can be superimposed onto each other or onto CAP with RMSD values ranging from 1.1 Å to 1.3 Å. Although the CAP-DNA structure (62) has revealed that 13 amino acid residues are involved in protein-DNA contacts, only three of them are base-specific: Arg180, Glu181, and Arg185. In the two superimposed HTHs, these residues correspond to Glu32, Thr33, and Lys37 of LrpA (Fig. 4).

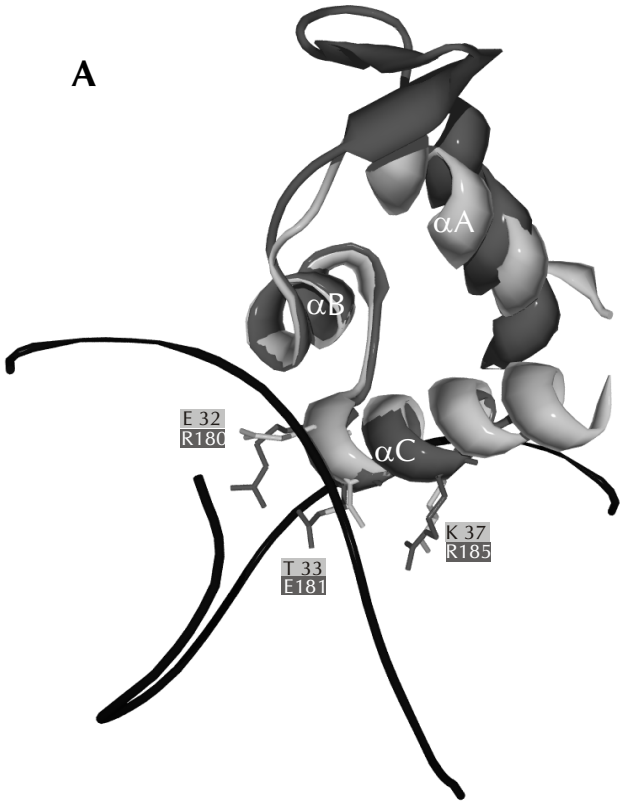
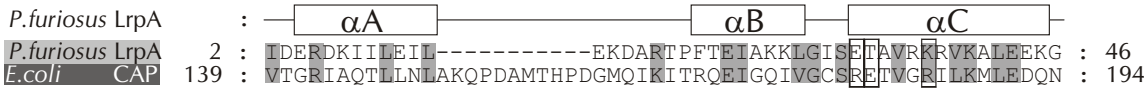



**Figure 4.** (**A**) Superimposition of the HTH domains of *P.furiosus* LrpA (36) and the *E.coli* CAP-DNA complex (62). White, LrpA HTH; gray, CAP HTH; black, DNA. αA, αB, and αC refer to the α-helices of the LrpA HTH (see also Fig. 3). The residues indicated are in CAP involved in base-specific CAP-DNA contacts, and predicted to have the same function in LrpA. (**B**) Alignment of the HTHs of CAP and LrpA. The three α-helices of the LrpA HTH are indicated by rectangles. Resdues involved in base-specific CAP-DNA interactions are boxed

**B**



Interestingly, as shown in the alignment of Figure 1A, the region containing these amino acids appears to be the most variable region of the HTH domain of Lrp-like proteins, possibly reflecting the variation in DNA-binding site specificity. It should be noted that Glu32 of LrpA corresponds to a proline (Pro41) in enterobacterial Lrp's and some other Lrp-like proteins. Because the side chain of a proline is not expected to be involved in base-specific DNA contacts, it is anticipated that protein-base contacts could possibly be somewhat different between the various Lrp-like proteins.

The HTH of LrpA is connected with a hinge to its C-terminal domain. In *E.coli* Lrp the latter domain has been implicated in the response to leucine and activation of transcription, as determined by a random mutational analysis of the protein (58). This C-terminal domain has connectivity βαββαβ (Fig. 5A), where the two α-helices are packed on one side of the four-stranded anti-parallel β-sheet. LrpA forms a homodimer mainly through interactions between the β-strands of the C-terminal domain, and an octamer through further interactions between the last α-helix and β-strand of the βαββαβ-fold (see Fig. 3). The C-terminal domain of Lrp-like proteins in general thus appears to be involved in ligand-response, activation, dimerization and multimerization. Possibly, interaction of amino acid ligands alters the multimerization properties of the protein, thereby affecting for example the cooperativity of DNA binding. An *E.coli* Lrp leucine-response mutant has been isolated in which DNA binding activity towards upstream *ilvIH* binding sites was not affected, however, cooperativity for binding to additional downstream sites was strongly reduced (58). This particular mutation corresponds to a residue located at the dimer interface of the C-terminus of *P.furiosus* LrpA. In addition, other leucine-response mutations mapped either within the dimer or octamer interface, suggesting that the action of amino acid ligands and multimerization are somehow related. This was recently demonstrated for *E.coli* Lrp, where multimerization was studied in the presence and absence of its ligand leucine (14). Leucine was found to promote dissociation of the Lrp hexadecamer into the octamer. Since these different oligomeric states of the protein are thought to prefer targets with a different binding site organization, it is conceivable that amino acid ligands modulate binding affinity through altering the oligomeric state of Lrp's (14). Unfortunately, no amino acid ligand has been identified yet for *P.furiosus* LrpA, so at present it is not possible to study the effect of ligand binding on the overall structure of LrpA.

## The RAM-domain

The C-terminal amino acid regulatory domain of LrpA is connected to the HTH domain by a flexible hinge. In a similar way ACT domains (Aspartokinase, Chorismate mutase and TyrA) are coupled to several proteins related to amino acid or purine metabolism (2). ACT domains are regulatory domains separated from the catalytic domains of their respective proteins by a flexible linker region. Allosteric regulation occurs as a consequence of the binding of an effector molecule (e.g. serine for SerA, the 3-phosphoglycerate dehydrogenase of *E.coli*) (27, 61) affecting enzyme activity through a slight interdomain rearrangement. ACT domains are coupled to proteins with enzymatic activity as well as DNA-binding activity (e.g. TyrR, regulator for tyrosine biosynthesis) (57).

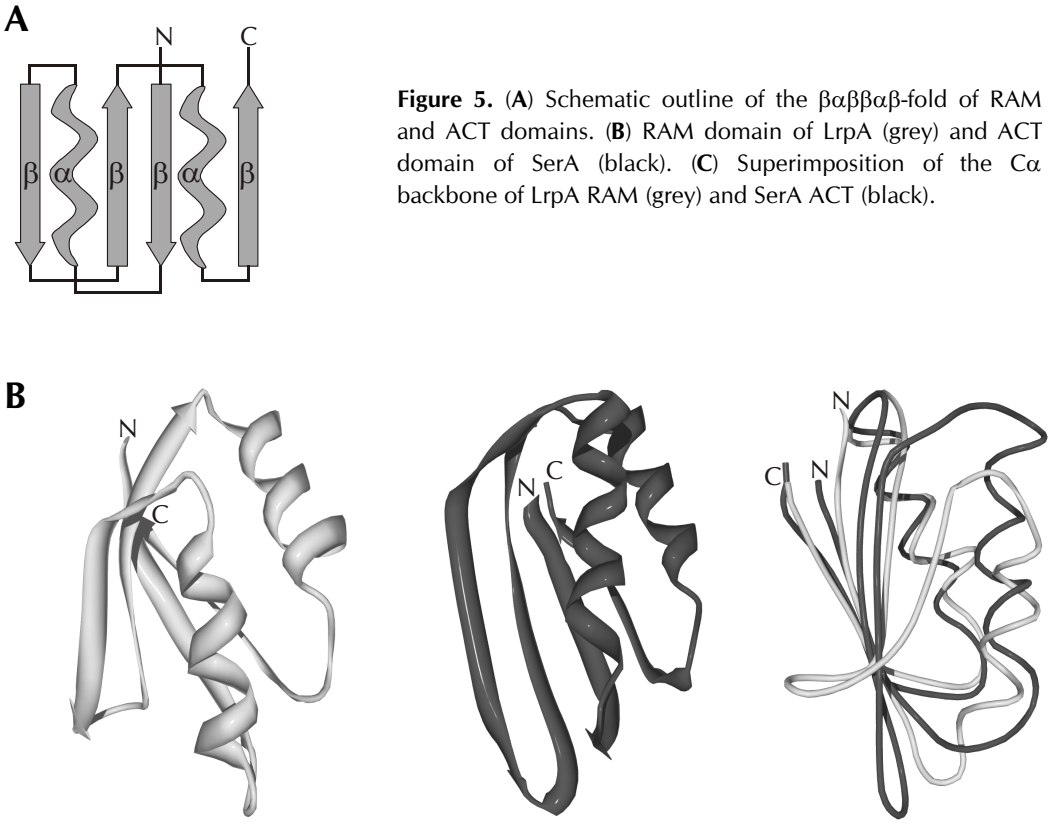



**Figure 5.** (**A**) Schematic outline of the βαββαβ-fold of RAM and ACT domains. (**B**) RAM domain of LrpA (grey) and ACT domain of SerA (black). (**C**) Superimposition of the Cα backbone of LrpA RAM (grey) and SerA ACT (black).

A comparison of ACT domains with the C-terminal domain of *P.furiosus* LrpA shows that although they share no significant homology at the amino acid sequence level, both domains have a comparable βαββαβ-fold. Superimposition of the C-terminal domain of LrpA and the ACT-domains of SerA and phenylalanine hydroxylase (PheOH) (32) confirms their structural resemblance (see Fig. 5) (23). As discussed below, the C-terminal domain of Lrp appears to be a common regulatory structure in amino acid metabolic enzymes and transcriptional regulators, and hence we recently proposed to refer to it as

RAM, after Regulator of Amino acid Metabolism (23). Because the sequence of the regulatory domain of Lrp-like proteins is not related to that of ACT domains, it should be regarded as a novel motif. An iterative database search (1) at the NCBI (National Center for Biotechnology Research) using the RAM-domain of *P.furiosus* LrpA, revealed an ubiquitous phyletic distribution of this domain among archaea and bacteria (about 170 proteins converged after 6 iterations (23)). The proteins that were retrieved using this search were prevailed by Lrp-like proteins, where an N-terminal HTH is fused to a C-terminal RAM domain (HTH-hinge-RAM). Additionally, examples were found of a RAM-domain that was fused to the C-terminus of 2-isopropylmalate synthase (IPM-RAM) in *S.solfataricus, S.tokodai*, and *Pyrobaculum aerophilum* (Fig. 6). IPM catalyzes the first step in leucine biosynthesis, and is anticipated to be subjected to RAM-domain mediated feedback regulation in these species. Besides, several isolated RAM-domains were identified in this PSI-BLAST search. This topological variability of the RAM-domain resembles that of the ACT domain, since ACT is also found as an allosteric regulatory domain associated with metabolic enzymes (like SerA, PheOH and TD (threonine deaminase) (26) and transcriptional regulators of amino acid metabolism like TyrR and PhhR (activator of the phenylalanine hydroxylase gene cluster) (63) (2). In addition, a stand-alone version of ACT is represented by IlvH, that acts as a regulatory subunit of the acetohydroxyacid synthase (IlvI), a key enzyme in branched chain amino acid biosynthesis that is subjected to valine feedback inhibition (47).

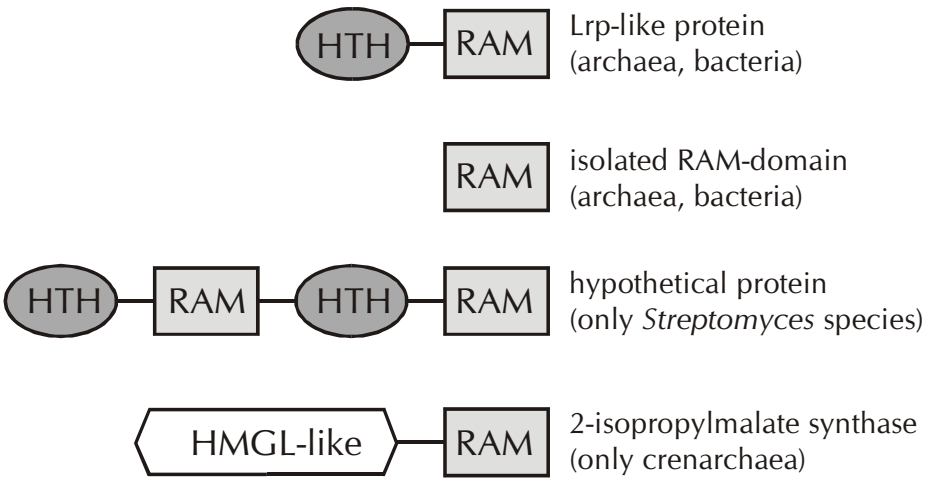



**Figure 6.** Topological variability of RAM domains, based on PSI-BLAST database analysis using the RAM-domain of *P.furiosus* LrpA (23).

It is anticipated that the several isolated archaeal RAM domains that were retrieved during the iterative database search, perform a function that is similar to that of IlvH: allosteric regulation of enzymes or transcriptional regulators involved in amino acid metabolism. In the case of TyrR the effect of the amino acid ligand tyrosine resembles the effect of leucine on Lrp, i.e. ligand-regulated self-association (3, 14). This may indicate that RAM and ACT domains function in a mechanistically similar way in these regulators. Apart from the apparent structural conservation, it should be noted that the expression of some bacterial ACT-containing enzymes (e.g. SerA, IlvH) is under control of Lrp-like (HTH-RAM) regulators (13). Although this suggests a phylogenetic link between Lrp, SerA, and IlvH, it should be noted that both structure-based alignments of the regulatory domains and extensive (reverse) PSI-BLAST analyses do not reveal any significant conservation of RAM and ACT at the protein sequence level. In addition, the ACT-domain is primarily found in association with metabolic enzymes, whereas the RAM domain is most often associated with transcriptional regulators, based on the currently available data. This suggests that the RAM and ACT domains are the result of convergent evolution, although their anticipated ligand-binding sites appear to be different (23).

It is concluded that both domains of Lrp-like proteins are independently present in other proteins as well. The HTH motif is a very common motif, responsible for specific DNA-binding activity, and the RAM-domain appears to be an allosteric regulatory domain that interacts with certain amino acid ligands. All current public databases contain several RAM-containing ORFs that are incorrectly annotated as Lrp-like proteins, since an N-terminal HTH is lacking. Likewise, a thorough analysis of the primary sequence of the previously annotated Lrp-like protein Lrs14 (Table I and Figure 1) (48) shows that although its N-terminal sequence clearly resembles a HTH motif, its C-terminal domain lacks significant homology with a RAM domain. Moreover, Lrs14 has the highest homology with non-Lrp-like DNA-binding proteins, especially from *Sulfolobales*, suggesting that this regulator is a novel type of (cren)archaeal-type regulator rather than an Lrp-like protein. To avoid annotation problems like this, we propose to define an Lrp-like protein as a protein with an N-terminal HTH domain connected to a C-terminal RAM domain.

## REFERENCES


1. Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-402.
2. Aravind, L., and E.V. Koonin 1999. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol 287:1023-40.
3. Bailey, M. F., B.E. Davidson, A.P. Minton, W.H. Sawyer, and G.J. Howlett 1996. The effect of self-association on the interaction of the *Escherichia coli* regulatory protein TyrR with DNA. J Mol Biol 263:671-84.
4. Barber, A.M., V.B. Zhurkin, and S. Adhya 1993. CRP-binding sites: evidence for two structural classes with 6-bp and 8- bp spacers. Gene 130:1-8.
5. Belitsky, B.R., Gustafsson, M.C., Sonenshein, A.L., Von Wachenfeldt, C. 1997. An *lrp*-like gene of *Bacillus subtilis* involved in branched-chain amino acid transport. J Bacteriol 179:5448-5457.
6. Bell, S.D., and S.P. Jackson 2001. Mechanism and regulation of transcription in archaea. Curr Opin Microbiol 4:208-13.
7. Bell, S.D., and S.P. Jackson 2000. Mechanism of autoregulation by an archaeal transcriptional repressor. J Biol Chem 275:31624-9.
8. Beloin, C., S. Ayora, R. Exley, L. Hirschbein, N. Ogasawara, Y. Kasahara, J.C. Alonso, and F. L. Hegarat 1997. Characterization of an *lrp*-like (*lrpC*) gene from *Bacillus subtilis*. Mol Gen Genet 256:63-71.
9. Beloin, C., R. Exley, A.L. Mahe, M. Zouine, S. Cubasch, and F. Le Hegarat 2000. Characterization of LrpC DNA-binding properties and regulation of *Bacillus subtilis lrpC* gene expression. J Bacteriol 182:4414-24.
10. Blattner, F.R., G. Plunkett, 3rd, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, and Y. Shao 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277:1453-74.
11. Brinkman, A. B., Bell, S.D., Lebbink, R.J., de Vos, W.M., van der Oost, J. 2002. The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *Submitted for publication.*
12. Brinkman, A.B., I. Dahlke, J.E. Tuininga, T. Lammers, V. Dumay, E. de Heus, J.H.G. Lebbink, M. Thomm, W.M. de Vos, J. van der Oost 2000. An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. J Biol Chem 275:38160-38169.
13. Calvo, J.M., and R.G. Matthews 1994. The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiol Rev 58:466-490.
14. Chen, S., M.H. Rosner, and J.M. Calvo 2001. Leucine-regulated self-association of leucine-responsive regulatory protein (Lrp) from *Escherichia coli*. J Mol Biol 312:625-35.
15. Cho, K., and S.C. Winans 1996. The *putA* gene of *Agrobacterium tumefaciens* is transcriptionally activated in response to proline by an Lrp-like protein and is not autoregulated. Mol Microbiol 22:1025-33.
16. Cui, Y., Wang, Q., Stormo, G.D., Calvo, J.M. 1995. A consensus sequence for binding of Lrp to DNA. J Bacteriol 177:4872-4880.

17. Dahlke, I., and M. Thomm 2002. A *Pyrococcus* homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment. Nucleic Acids Res 30:701-10.
18. Dartois, V., Lui, J., Hoch, J.A. 1997. Alterations in the flow of one-carbon units affect KinB-dependent sporulation in *Bacillus subtilis*. Mol Microbiol 25:39-51.
19. de Wind, N., M. de Jong, M. Meijer, and A.R. Stuitje 1985. Site-directed mutagenesis of the *Escherichia coli* chromosome near *oriC*: identification and characterization of *asnC*, a regulatory element in *E.coli* asparagine metabolism. Nucleic Acids Res 13:8797-811.
20. Enoru-Eta, J., D. Gigot, T.L. Thia-Toong, N. Glansdorff, and D. Charlier 2000. Purification and characterization of Sa-Lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon *Sulfolobus acidocaldarius* homologous to the bacterial global transcriptional regulator Lrp. J Bacteriol 182:3661-72.
21. Ernsting, B.R., M.R. Atkinson, A.J. Ninfa, and R.G. Matthews 1992. Characterization of the regulon controlled by the leucine-responsive regulatory protein in *Escherichia coli*. J Bacteriol 174:1109-18.
22. Ernsting, B.R., J.W. Denninger, R.M. Blumenthal, and R.G. Matthews 1993. Regulation of the *gltBDF* operon of *Escherichia coli*: how is a leucine-insensitive operon regulated by the leucine-responsive regulatory protein? J Bacteriol 175:7160-9.
23. Ettema, T.J.G., A.B. Brinkman, P.M. Leonard, T.H. Tani, J.B. Rafferty, and J. van der Oost 2002. A novel ligand-binding domain (RAM) involved in allosteric regulation of amino acid metabolism in prokaryotes. J Biol Chem, *in press*.
24. Friedberg, D., M. Midkiff, and J.M. Calvo 2001. Global versus local regulatory roles for Lrp-related proteins: *Haemophilus influenzae* as a case study. J Bacteriol 183:4004-11.
25. Friedberg, D., Platko, J.V., Tyler, B., Calvo, J.M. 1995. The amino acid sequence of Lrp is highly conserved in four enteric microorganisms. J Bacteriol 177:1624-1626.
26. Gallagher, D.T., G.L. Gilliland, G. Xiao, J. Zondlo, K.E. Fisher, D. Chinchilla, and E. Eisenstein 1998. Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase. Structure 6:465-75.
27. Grant, G.A., D.J. Schuller, and L.J. Banaszak 1996. A model for the regulation of D-3-phosphoglycerate dehydrogenase, a $V_{max}$-type allosteric enzyme. Protein Sci 5:34-41.
28. Hay, N.A., D.J. Tipper, D. Gygi, and C. Hughes 1997. A nonswarming mutant of *Proteus mirabilis* lacks the Lrp global transcriptional regulator. J Bacteriol 179:4741-6.
29. Inoue, H., K. Inagaki, S.I. Eriguchi, T. Tamura, N. Esaki, K. Soda, and H. Tanaka 1997. Molecular characterization of the *mde* operon involved in L-methionine catabolism of *Pseudomonas putida*. J Bacteriol 179:3956-62.
30. Jafri, S., Evoy, S., Cho, K., Craighead, H.G., 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.
31. Keuntje, B., Masepohl, B., 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.
32. Kobe, B., I.G. Jennings, C.M. House, B.J. Michell, K.E. Goodwill, B.D. Santarsiero, R.C. Stevens, R.G. Cotton, and B.E. Kemp 1999. Structural basis of autoregulation of phenylalanine hydroxylase. Nat Struct Biol 6:442-8.

33. Kölling, R., A. Gielow, W. Seufert, C. Kucherer, and W. Messer 1988. AsnC, a multifunctional regulator of genes located around the replication origin of *Escherichia coli*, *oriC*. Mol Gen Genet 212:99-104.
34. Kölling, R., and H. Lother 1985. AsnC: an autogenously regulated activator of asparagine synthetase A transcription in *Escherichia coli*. J Bacteriol 164:310-5.
35. Lai, H.C., D. Gygi, G.M. Fraser, and C. Hughes 1998. A swarming-defective mutant of *Proteus mirabilis* lacking a putative cation-transporting membrane P-type ATPase. Microbiology 144:1957-61.
36. Leonard, P.M., S.H.J. Smits, S.E. Sedelnikova, A.B. Brinkman, W.M. de Vos, J. van der Oost, D.W. Rice, and J.B. Rafferty 2000. Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. EMBO J 20:990-997.
37. Levinthal, M., P. Lejeune, and A. Danchin 1994. The H-NS protein modulates the activation of the *ilvIH* operon of *Escherichia coli* K12 by Lrp, the leucine regulatory protein. Mol Gen Genet 242:736-43.
38. Lin, R., R. D'Ari, and E.B. Newman 1992. Lambda P $_{lac}$Mu insertions in genes of the leucine regulon: extension of the regulon to genes not regulated by leucine. J Bacteriol 174:1948-55.
39. Madhusudhan, K.T., Hester, K. L., Friend, V., Sokatch, J. R. 1997. Transcriptional activation of the *bkd* operon of *Pseudomonas putida* by BkdR. J Bacteriol 179:1992-1997.
40. Madhusudhan, K.T., N. Huang, E.H. Braswell, and J.R. Sokatch 1997. Binding of L-branched-chain amino acids causes a conformational change in BkdR. J Bacteriol 179:276-9.
41. Madhusudhan, K.T., N. Huang, J.R. Sokatch 1995. Characterization of BkdR-DNA binding in the expression of the *bkd* operon of *Pseudomonas putida*. J Bacteriol 177:636-641.
42. Madhusudhan, K.T., D. Lorenz, and J.R. Sokatch 1993. The *bkd*R gene of *Pseudomonas putida* is required for expression of the *bkd* operon and encodes a protein related to Lrp of *Escherichia coli*. J Bacteriol 175:3934-40.
43. Madhusudhan, K.T., J. Luo, and J.R. Sokatch 1999. *In vitro* transcriptional studies of the *bkd* operon of *Pseudomonas putida*: L-branched-chain amino acids and D-leucine are the inducers. J Bacteriol 181:2889-94.
44. Marshall, D.G., B.J. Sheehan, and C.J. Dorman 1999. A role for the leucine-responsive regulatory protein and integration host factor in the regulation of the *Salmonella* plasmid virulence (*spv*) locus in *Salmonella typhim*urium. Mol Microbiol 34:134-45.
45. Mathew, E., J. Zhi, M. Freundlich 1996. Lrp is a direct repressor of the *dad* operon in *Escherichia coli*. J Bacteriol 178:7234-7240.
46. McKay, D.B., and T.A. Steitz 1981. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. Nature 290:744-9.
47. Mendel, S., T. Elkayam, C. Sella, V. Vinogradov, M. Vya zmensky, D.M. Chipman, and Z. Barak 2001. Acetohydroxyacid synthase: a proposed structure for regulatory subunits supported by evidence from mutagenesis. J Mol Biol 307:465-77.
48. Napoli, A., J. van der Oost, C.W. Sensen, R.L. Charlebois, M. Rossi, and M. Ciaramella 1999. An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. J Bacteriol 181:1474-1480.
49. Newman, E.B., R. D'Ari, R.T. Lin 1992. The leucine-Lrp regulon in *E.coli*: a global response in search of a raison d'etre. Cell 68:617-619.
50. Newman, E.B., R. Lin 1995. Leucine-responsive regulatory protein: a global regulator of gene expression in *E.coli*. Ann Rev Microbiol 49:747-775.

51. Nicholson, B., and D.A. Low 2000. DNA methylation-dependent regulation of *pef* expression in *Salmonella typhimurium*. Mol Microbiol 35:728-42.
52. Nou, X., B.A. Braaten, L. Kaltenbach, D.A. Low 1995. Differential binding of Lrp to two sets of *pap* DNA binding sites mediated by PapI regulates Pap phase variation in *Escherichia coli*. EMBO J. 14:5785-5797.
53. Ouhammouch, M., and E.P. Geiduschek 2001. A thermostable platform for transcriptional regulation: the DNA-binding properties of two Lrp homologs from the hyperthermophilic archaeon *Methanococcus jannaschii*. EMBO J 20:146-56.
54. Paul, L., R.M. Blumenthal, and R.G. Matthews 2001. Activation from a distance: roles of Lrp and integration host factor in transcriptional activation of *gltBDF*. J Bacteriol 183:3910-8.
55. Peekhaus, N., and R. Kramer 1996. The *gluEMP* operon from *Zymomonas mobilis* encodes a high-affinity glutamate carrier with similarity to binding-protein-dependent transport systems. Arch Microbiol 165:325-32.
56. Peekhaus, N., B. Tolner, B. Poolman, R. Kramer 1995. The glutamate uptake regulatory protein (Grp ) of *Zymomonas mobilis* and its relation to the global regulator Lrp of *Escherichia coli*. J Bacteriol 177:5140-5147.
57. Pittard, A.J., and B.E. Davidson 1991. TyrR protein of *Escherichia coli* and its role as repressor and activator. Mol Microbiol 5:1585-92.
58. Platko, J.V., J.M. Calvo 1993. Mutations affecting the ability of *Escherichia coli* Lrp to bind DNA, activate transcription, or respond to leucine. J Bacteriol 175:1110-1117.
59. Roesch, P.L., and I.C. Blomfield 1998. Leucine alters the interaction of the leucine-responsive regulatory protein (Lrp) with the *fim* switch to stimulate site-specific recombination in *Escherichia coli*. Mol Microbiol 27:751-61.
60. Schevitz, R.W., Z. Otwinowski, A. Joachimiak, C.L. Lawson, and P.B. Sigler 1985. The three-dimensional structure of Trp repressor. Nature 317:782-6.
61. Schuller, D.J., G.A. Grant, and L.J. Banaszak 1995. The allosteric ligand site in the $V_{max}$-type cooperative enzyme phosphoglycerate dehydrogenase. Nat Struct Biol 2:69-76.
62. Schultz, S.C., G.C. Shields, T.A. Steitz 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science 253:1001-1007.
63. Song, J., and R.A. Jensen 1996. PhhR, a divergently transcribed activator of the phenylalanine hydroxylase gene cluster of *Pseudomonas aeruginosa*. Mol Microbiol 22:497-507.
64. Travers, A.A. 1989. DNA conformation and protein binding. Annual Rev Biochem 58:427-452.
65. Tuerk, C., and L. Gold 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505-10.
66. van der Woude, M.W., and D.A. Low 1994. Leucine-responsive regulatory protein and deoxyadenosine methylase control the phase variation and expression of the *sfa* and *daa* pili operons in *Escherichia coli*. Mol Microbiol 11:605-18.
67. Wang, Q., J.M. Calvo 1993. Lrp, a global regulatory protein of *Escherichia coli*, binds cooperatively to multiple sites and activates transcription of *ilvIH*. J Mol Biol 229:306-318.
68. Wang, Q., J.M. Calvo 1993. Lrp, a major regulatory protein in *Escherichia coli*, bends DNA and can organize the assembly of a higher-order nucleoprotein structure. EMBO J 12:2495-2501.

69. Wang, Q., M. Sacco, E. Ricca, C.T. Lago, M. De Felice, J.M. Calvo 1993. Organization of Lrp -binding sites upstream of *ilvIH* in *Salmonella typhimurium*. Mol Microbiol 7:883-891.
70. Weyand, N.J., B.A. Braaten, M. van der Woude, J. Tucker, and D.A. Low 2001. The essential role of the promoter-proximal subunit of CAP in Pap phase variation: Lrp - and helical phase-dependent activation of *papBA* transcription by CAP from -215. Mol Microbiol 39:1504-22.
71. Weyand, NJ., and D.A. Low 2000. Regulation of Pap phase variation. Lrp is sufficient for the establishment of the phase off *pap* DNA methylation pattern and repression of *pap* transcription *in vitro*. J Biol Chem 275:3192-200.
72. Wiese, D.E. 2nd., B.R. Ernsting, R.M. Blumenthal, R.G. Matthews 1997. A nucleoprotein activation complex between the leucine-responsive regulatory protein and DNA upstream of the *gltBDF* operon in *Escherichia coli*. J Mol Biol 270:152-168.
73. Wintjens, R., and M. Rooman 1996. Structural classification of HTH DNA -binding domains and protein-DNA interaction modes. J Mol Biol 262:294 -313.
74. Zeng, Q., and A.O. Summers 1997. A glutamate uptake regulatory protein (G rp) in *Escherichia coli*? Mol Microbiol 24:231-2.

# Chapter 4

## An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated

Arie B. Brinkman
Isabell Dahlke
Judith E. Tuininga
Torsten Lammers
Valerie Dumay
Edwin de Heus
Joyce H.G. Lebbink
Michael Thomm
Willem M. de Vos
John van der Oost

**The archaeal transcriptional initiation machinery closely resembles core elements of the eukaryal polymerase II system. However, apart from the established basal archaeal transcription system, little is known about the modulation of gene expression in archaea. At present, no obvious eukaryal-like transcriptional regulators have been identified in archaea. Instead, we have previously isolated an archaeal gene, *Pyrococcus furiosus lrpA*, which potentially encodes a bacterial-like transcriptional regulator. In the present study, we have for the first time addressed the actual involvement of an archaeal Lrp homologue in transcription modulation. For that purpose, we have produced LrpA in *Escherichia coli*. In a cell-free *P. furiosus* transcription system we used wild-type and mutated *lrpA* promoter fragments to demonstrate that the purified LrpA negatively regulates its own transcription. In addition, gel retardation analyses revealed a single protein-DNA complex, in which LrpA appeared to be present in (at least) a tetrameric conformation. The location of the LrpA binding site was further identified by DNaseI and hydroxyl radical footprinting, indicating that LrpA binds to a 46-bp sequence that overlaps the transcriptional start site of its own promoter. The molecular basis of the transcription inhibition by LrpA is discussed.**

*Journal of Biological Chemistry, 2000, 275 (49): 38160-38169*

## INTRODUCTION

Recent studies have revealed that the archaeal transcriptional machinery represents a simplified version of the eukaryal RNA polymerase II transcription apparatus, which involves homologues of the TATA binding protein (TBP), the transcription factor IIB (TFIIB, the archaeal homologue is called TFB) and the multi-subunit RNA polymerase II (for a recent review, see (2)). The initiation process starts when the TBP interacts specifically with the core promoter element, the TATA box, which is located at positions -25 to -30 relative to the transcriptional start site (+1). This complex is stabilized by TFB, which interacts with TBP as well as with the nucleotides -42 to -19 that flank the TATA box (23). In particular, a sequence upstream of the TATA-box (called the TFB-responsive element or BRE) is essential for transcriptional polarity (3, 39). Formation of this pre-initiation complex results in recruitment of the RNA polymerase complex (2). Although important progress has recently been made with the elucidation of the archaeal transcriptional mechanism, very little is yet known about the actual regulation of this process. A limited number of studies reported that expression of genes involved in nitrogen metabolism, methanogenesis, and sugar metabolism are subject to substrate-dependent regulation at the transcriptional level (9, 10, 42, 45, 62). Unfortunately, data about the molecular mechanisms underlying this regulation are still scarce. One of the few transcriptional regulators that have recently been studied in more detail concerns GvpE, an activator that is required for the expression of genes involved in gas vesicle synthesis in halophilic Archaea. In a molecular modeling study, GvpE has been proposed to resemble a eukaryal leucine-zipper dimer, that might interact with a palindrome sequence of its target promoter, centered 40-50 bp upstream of the transcriptional start site (35). Another putative transcriptional regulator that has been studied in more detail is Tfx from *Methanobacterium thermoautotrophicum* (26). The *tfx*-encoding gene is located upstream of the operon encoding molybdenum formylmethanofuran dehydrogenase (*fmdECB*). Tfx binds to a site located 167 bp downstream of the transcriptional start site of *fmdE*. It was proposed that Tfx is a transcriptional activator required for the expression of *fmdECB*. Obvious homologues of Tfx can only be found within the domain of the archaea.

Analysis of the available archaeal genomic sequences shows that the majority of the identified homologues of regulators are bacterial-like (37). Recently a mechanism by which a bacterial-like regulator affects the archaeal transcriptional machinery was described. It was shown that MDR1 from *Archaeoglobus fulgidus*, a homologue of the iron-dependent bacterial repressor DxtR, represses transcription by binding to its own promoter in a metal-

dependent manner. Upon binding of MDR1 to the promoter, RNA polymerase recruitment is prevented, but not binding of TBP or TFB (1).

One particular group of bacterial-like regulators present in all available archaeal genomes is the family of Lrp/AsnC regulators. Members of this family have been identified in over a dozen different bacterial species, in which they generally appear to be involved in regulation of amino acid metabolism. The most extensively studied example is the leucine-responsive regulatory protein (Lrp) from *Escherichia coli* (5, 44). Lrp is a global regulator that controls the expression of approximately 75 genes, many of which are involved in transport, degradation or biosynthesis of amino acids. Lrp can either activate or repress transcription, and this action can be modulated by the effector leucine, which either decreases or increases its particular action. In some cases, like the negative autoregulation, leucine has no effect at all. The paralogous *E.coli* AsnC appears to be a specific transcriptional activator of asparagine synthetase A. Activation is reduced in the presence of the effector asparagine, but again, the negative autoregulation of AsnC itself is not affected by asparagine (13, 34). Two Lrp/AsnC homologues from *Sulfolobus solfataricus* have been studied. One of these was cloned, sequenced, and shown to be expressed during growth on complex medium (8). The other Lrp/AsnC homologue, called Lrs14, was studied in more detail (43). It was shown that the purified recombinant protein binds to its promoter at a region overlapping the TATA box. In addition, it was shown that the *lrs14* transcript accumulates in the late growth stages of *S.solfataricus*.

In the genome sequence of *Pyrococcus furiosus* (http://comb5-156.umbi.umd.edu/genemate/), at least 10 homologues of genes encoding Lrp/AsnC-like proteins can be identified (Brinkman and van der Oost, unpublished). The *lrpA* gene, encoding one of these homologues, was previously identified downstream of the *gdh* gene that encodes glutamate dehydrogenase (15, 36). In this paper, we describe the cloning, functional expression, and characterization of LrpA, and show that LrpA binds to its own promoter and specifically inhibits *in vitro* transcription from this promoter. Using the combined data of gel-mobility shift assays, *in vitro* transcription analyses, and footprinting, we identified the sequence elements responsible for LrpA binding, and propose a mechanism by which LrpA binds its promoter.

# *RESULTS*

## LrpA sequence analysis

A 2.7-kb *Hin*dIII fragment including the gene encoding glutamate dehydrogenase (*gdh*) was previously isolated from a genomic library of *P.furiosus* (Fig. 1B) (15). Downstream of the *gdh* gene, an open reading frame was found with a high degree of similarity to bacterial transcription regulators of the Lrp/AsnC family (36). Sequence analysis of the *lrpA* gene

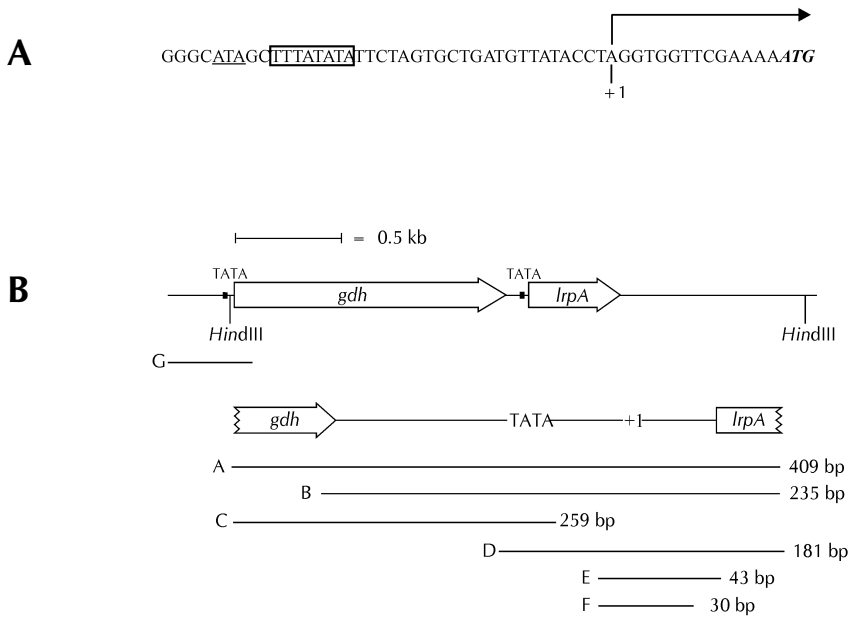



**Figure 1.** (**A**) Sequence of the *lrpA* promoter. The transcriptional start (+1) is indicated with a horizontal arrow. Underlined, BRE-element; boxed, TATA-element; bold Italics, ATG-startcodon of *lrpA*. (**B**) The *gdh-lrpA* locus and DNA fragments used in this study. *Hin*dIII restriction sites indicate the genomic fragment carrying the glutamate dehydrogenase-encoding gene (*gdh*) and the gene encoding the Lrp-like regulator (LrpA). Filled squares show the TATA boxes of the *gdh* and *lrpA* promoters. Dashed lines display the enlarged intergenic region between *gdh* and *lrpA* with the TATA boxes of the *lrpA* promoter, the *lrpA* transcriptional start site (+1), and DNA fragments (A-G) used in experiments.

identified a frameshift in the previously published sequence (accession number P42180), which introduced a stop codon after lysine 120 in the predicted protein sequence. The corrected *P.furiosus lrpA* gene is predicted to encode a 141 amino acid protein with a predicted molecular mass of 15.9 kDa. Subsequent BLAST analysis revealed that LrpA shares a high degree of similarity with many (hypothetical) regulatory proteins from a number of Archaea including *Pyrococcus horikoshii* (93% identity), *Pyrococcus abyssi* (98% identity), *Methanococcus jannaschii* (54% identity), *Archaeoglobus fulgidus* (49% identity),

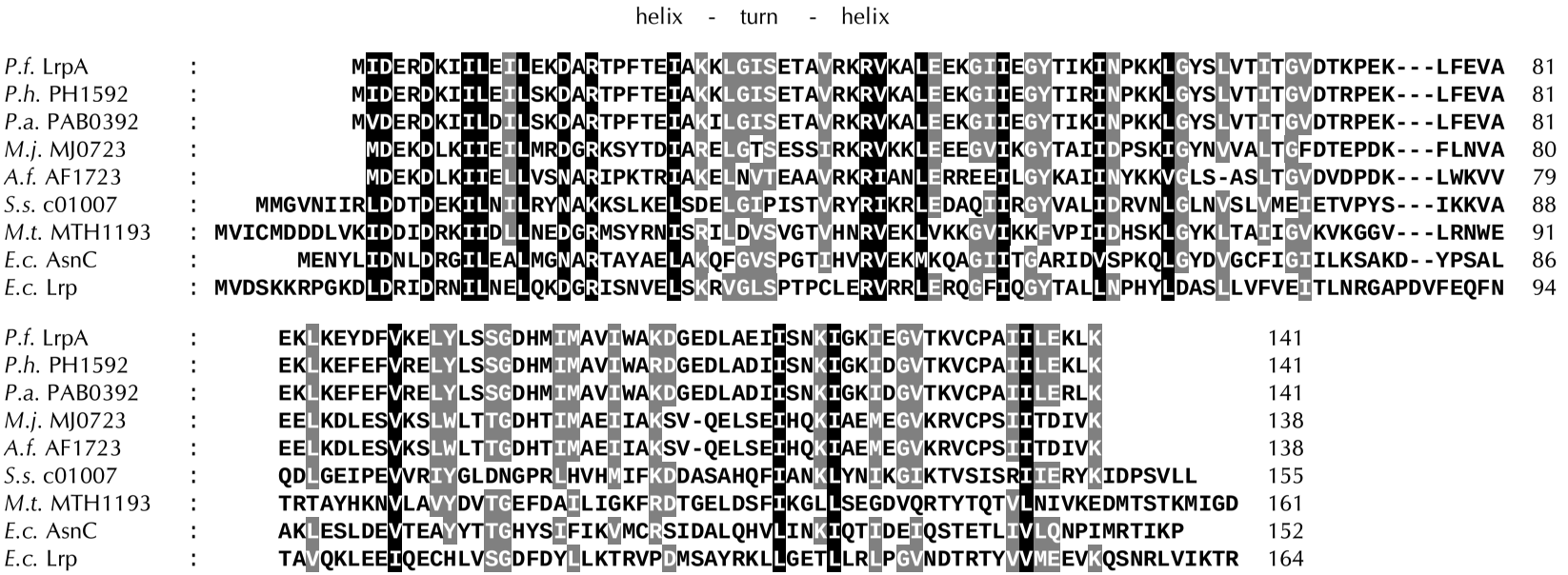



**Figure 2.** Alignment of *P.furiosus* LrpA with (putative) archaeal and bacterial Lrp/AsnC homologues. Homologues of the Lrp/AsnC family of transcriptional regulators can be identified in all sequenced archaeal genomes. From each sequenced archaeon the deduced amino acid sequence of a putative Lrp/AsnC-like regulator with highest homology to *P.furiosus* LrpA is aligned with that of *P.furiosus* LrpA, using the program ClustalX. In addition, the amino acid sequence of Lrp and AsnC from *E.coli* are aligned. Amino acids were grouped in four clusters: [I, L, V, M]; [R, K]; [F, Y, W] and [E, D, N, Q]. Black shaded symbols indicate completely conserved amino acid residues or amino acids belonging to the same cluster. Grey shaded symbols indicate seven or eight residues belonging to one cluster. The helix-turn-helix DNA-binding motif is boxed. *P.f.* LrpA, *P.furiosus* accession number P42180; *P.h.* PH1592, *P.horikoshii* accession number AP000006, 93% identity (132/141); *P.a.* PAB0392, *P.abyssi* accession number CAB49492, 98% identity (139/141); *M.j.* MJ0723, *M.jannaschii* accession number Q58133, 54% identity (77/141); *A.f.* AF1723, *A.fulgidus* accession number AE000984, 49% identity (69/140); *M.t.* MTH1193, *M.thermoautotrophicum* accession number AE000887, 37% identity (27/72), *S.s.* c01007, *S.solfataricus* accession number Y08256, 32% identity (46/141); *E.c.* AsnC, *E.coli* accession number P03809, 33% identity (46/139); *E.c.* Lrp, *E.coli* accession number P19494, 28% identity (41/143).

*Methanobacterium thermoautotrophicum* (37% identity), and *Sulfolobus solfataricus* (32% identity, Fig. 2). In addition, all of these Archaea contain a number of more distantly related homologues, e.g. in *P.horikoshii* a total of 9 genes appear to encode LrpA homologues, while *P.furiosus* itself contains at least 10 LrpA homologues. The best characterized LrpA homologues are from bacterial origin, in particular *E.coli* AsnC (33% identity) and *E.coli* Lrp (28% identity). A PROSITE pattern search with *P.furiosus* LrpA identified a putative helix-turn-helix motif of the Lrp/AsnC family (Fig. 2). This motif was also predicted by the program HELIX-TURN-HELIX (14).

## Overexpression and purification of LrpA

Using a PCR approach we cloned the *lrpA* gene into a pET9d vector, resulting in pLUW604. Production of LrpA was achieved after transformation of pLUW604 to *E.coli* BL21(λDE3). After O/N growth in the presence of 0.4 mM IPTG, cells were harvested and disrupted. Tricine-SDS-PAGE analysis of membrane and soluble fractions indicated that 50% percent of the produced LrpA was present as soluble protein (Fig. 3, lane 3 and 4). The soluble fraction containing LrpA was further subjected to a heat incubation of 30 min at 80°C, resulting in the denaturation of most of the *E.coli* proteins (Fig. 3, lane 5).

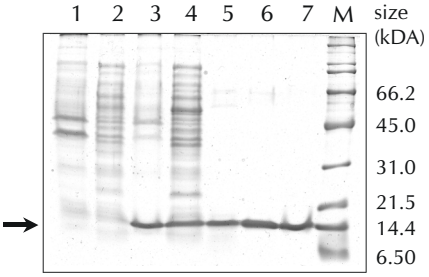



**Figure 3.** Overexpression and purification of LrpA. 10% Tricine-SDS-PAGE (52) showing extracts of *E.coli* BL21(DE3) after several steps of purification. Lane 1 and 2, insoluble and soluble fraction of extract from pET9d-containing strain, respectively; lane 3 and 4, insoluble and soluble fraction of extract from pLUW604-containing strain, respectively. These extracts clearly show overexpression of *P.furiosus* LrpA (arrow). The extract shown in lane 4 was subjected to a 30′ heat-treatment at 80°C and was subsequently centrifuged to separate denatured proteins from heat-stable proteins, resulting in a heat-stable cell free extract (HSCFE, lane 5). This HSCFE was used for further purification, consisting of cation-exchange chromatography (lane 6) and gel filtration (lane 7).

This heat-stable cell free extract was used for further purification by cation-exchange chromatography and gel filtration chromatography (Fig. 3, lanes 6 and 7). The calculated molecular mass of LrpA is 15.9 kDa, which is in good agreement with its migration on SDS-PAGE (Fig. 3). Elution patterns from gel filtration showed peaks corresponding to molecular masses of approximately 30, 60, and 120 kDa. This suggests that LrpA exists as a dimer, tetramer, and octamer in solution. We performed several independent gel filtration experiments with LrpA, and the apparent oligomeric heterogeneity was always observed.

## Analysis of the *lrpA* promoter

We used primer extension analysis to map the transcriptional start site for *lrpA*. Using *in vitro* generated run-off transcript RNA (see below) we found that the transcriptional start was located at an adenosine, located 14 bp upstream the translational start (see Fig. 1A). We compared the *lrpA* promoter sequence to other known promoter sequences from *P.furiosus*. Although only 14 *Pyrococcus* promoters have been mapped to date, a clear consensus sequence can be derived for the *Pyrococcus* BRE and TATA-elements: AAAnnTTTWWWWW (-35 to -23 sequence relative to the transcriptional start (+1), where n = any base, and W = A or T). The putative BRE and TATA-elements of the *lrpA* promoter match well with the consensus sequence mentioned above (see Fig. 1A).

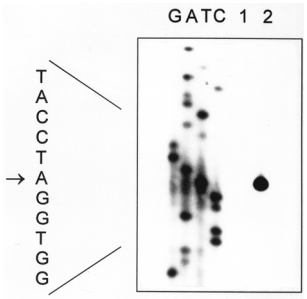



**Figure 4.** Analysis of the transcription start site of the *lrpA* promoter *in vitro* by primer extension. The sequence of the template DNA strand is shown the left of the primer extension product (marked G, A, T, C). Lane 1, control experiment (nucleotides omitted from transcription reaction); lane 2, analysis of the primer extension product.

We used total isolated RNA from *P.furiosus* grown on cellobiose, pyruvate, and tryptone to determine the transcriptional start *in vivo*. In all cases the transcriptional start was identical to that found with *in vitro* generated RNA (not shown), however, relatively weak signals were obtained. Although the results exhibited that *lrpA* is expressed during growth on above-mentioned substrates, they indicate that *lrpA* transcript levels are not very abundant under these conditions.

## *In vitro* transcription of *lrpA* and *gdh* in the presence of LrpA

Several bacterial members of the Lrp/AsnC family are negatively autoregulated (32, 34, 65). Hence, it was anticipated that *P.furiosus* LrpA could also be a repressor of its own expression. In the previously established cell-free transcription system (24), the *P.furiosus* purified transcription factors TBP, TFB and RNA polymerase direct efficient transcription from a DNA template containing the (partial) *gdh*-encoding sequence and its promoter. We used this *in vitro* transcription system to study the effect of LrpA on *lrpA* and *gdh* transcription (Fig. 5). The template used in this experiment is *Pst*I linearized pLUW613, which carried fragment A (Fig. 1B). Transcription from this template results in a 160-

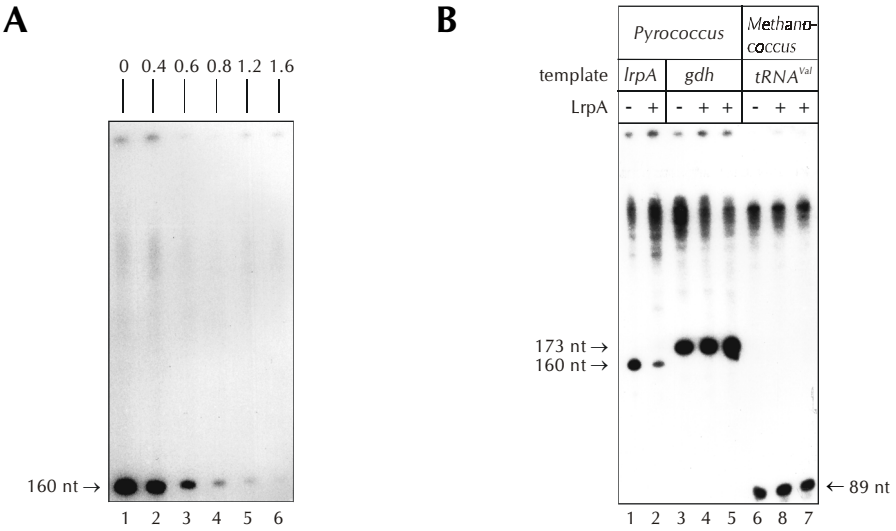



**Figure 5.** (**A**) Inhibition of *lrpA* transcription by LrpA. Linearized pLUW613 DNA containing the *P.furiosus lrpA* promoter was used as template in cell-free transcription experiments. The 160-nt run-off transcripts are indicated by an arrow. (**B**) Effects of LrpA on cell-free transcription from the *lrpA* and *gdh* promoter of *Pyrococcus* and the *tRNA*$^{val}$ promoter of *Methanococcus*. Transcripts from the *M.vannielii* promoter were analysed in the *in vitro* transcription system of *Pyrococcus* as shown previously (24). Addition of recombinant LrpA (0.8 μg) is indicated (+/-). The run-off transcripts are indicated by arrows, and their respective sizes are shown.

nucleotide *lrpA* run-off transcript. With increasing amounts of LrpA present in the reactions, the signal of the radiolabeled *lrpA Pst*I run-off transcripts drastically decreased (Fig. 5A). This indicated that LrpA had a negative effect on its own transcription *in vitro*. A similar experiment was performed using *Bam*HI-digested pLUW479 as DNA template (24). This plasmid carried the partial *P.furiosus gdh* gene including its 200-bp upstream sequence. Transcription from this template resulted in a 173-nucleotide *gdh* run-off transcript. There was no effect on *gdh* transcription when LrpA was added (Fig. 5B). Likewise, transcription from the *Methanococcus vannielii* $tRNA^{Val}$ promoter (24) was not inhibited by LrpA (Fig. 5B).

## LrpA binds specifically to its own promoter

To test whether LrpA binds to its own promoter, we performed gel-mobility shift experiments with purified LrpA and DNA fragments containing the LrpA promoter sequence. Three different DNA fragments (C, D, and E) were used, that contained sequences upstream of *lrpA* (Fig. 1B). Fragment C contained the upstream *lrpA* promoter region including the TATA element (Fig. 1B). LrpA did not shift this fragment, indicating that no interaction occurred between LrpA and the upstream promoter region (Fig. 6). Fragment D contained the TATA element, the transcriptional start site, and the sequence downstream thereof (Fig. 1B). Addition of LrpA to this fragment resulted in a shifted band

(Fig. 6). Using LrpA concentrations of up to 810 nM (as calculated for an LrpA monomer), we observed only one shifted band in gel-mobility shift experiments, indicating that only one complex is formed when LrpA binds to the *lrpA* promoter. This binding was specific, since the addition of 2 μg poly(dI.dC).poly(dI.dC) as nonspecific competitor DNA did not prevent LrpA-DNA binding.

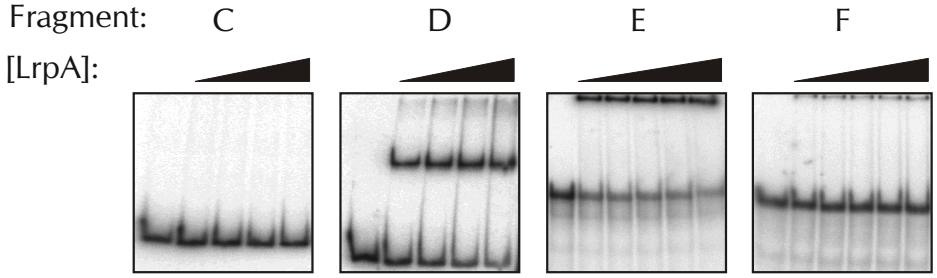



**Figure 6.** LrpA binds specifically to *lrpA* promoter DNA. 0, 30, 90, 270, or 810 nM of purified LrpA was added to fragments C and D. 2 μg of poly(dI.dC).poly(dI.dC) was present in these reactions. 0, 10, 25, 50, 100, or 200 nM of purified LrpA was added to fragments E and F. No aspecific competitor DNA was present in these reactions.

As mentioned above, we observed dimer, tetramer, and octamer configurations of LrpA in gel filtration chromatography experiments. We tested fractions of all these different forms in gel-mobility shift experiments, however, there was no obvious difference in DNA-binding activity between the fractions (not shown).

To determine the affinity of LrpA for its promoter DNA, we performed a gel-mobility shift experiment with fragment D and increasing LrpA concentrations (not shown). The concentration of LrpA, as calculated for an LrpA monomer, that caused half of the DNA to become complexed under the experimental conditions used was taken as the dissociation constant ($K_d$) of the LrpA-DNA complex (6). We determined a $K_d$ of 0.3 nM, which is about four to ten-fold lower compared to values measured for *E.coli* Lrp (48, 63), and several orders of magnitude lower than that of *S.solfataricus* Lrs14 (43).

We tested several smaller fragments to locate the boundaries of the LrpA binding sequence more precisely. Fragment E (43 bp) was the smallest fragment to which LrpA bound efficiently (Fig. 6 and 7C). LrpA also bound to fragments smaller than 43 bp, such as fragment F (30 bp), but the affinity for these fragments was drastically decreased (Fig. 6 and 7C). LrpA binding to fragment E was specific, since addition of increasing amounts of unlabelled fragment E prevented LrpA binding to labelled fragment D (not shown). Altogether, these results indicated that LrpA bound specifically to its own promoter, at a position around the transcriptional start site. Since *P.furiosus* grows optimally at 100°C we

performed binding reactions for gel-mobility experiments at several temperatures. Binding experiments performed at 0°C, 4°C, 25°C, 50°C, and 80°C resulted in identical gel-mobility shift patterns (not shown).

Gel-mobility shift experiments were also performed with DNA fragments containing *gdh* promoter sequences. Under similar conditions as the experiments with the *lrpA* promoter fragments, we observed very weak LrpA binding with fragment G (not shown). This fragment contained 386 bp of the *gdh* promoter, including the TATA element (Fig. 1B). However, this binding appeared to be rather weak and nonspecific since the observed shift disappeared upon the addition of poly(dI.dC).poly(dI.dC). This indicates that, under the tested conditions, LrpA binds specifically to its own promoter, but not to the *gdh* promoter. With promoter fragments of both *lrpA* and *gdh* several compounds were tested for their ability to act as an effector for LrpA. Among those tested were L-leucine, L-alanine, L-glutamate, α-ketoglutarate, pyruvate, cellobiose, and ammonium. However, none of these compounds affected LrpA binding at either the *lrpA* or the *gdh* promoter.

## DNaseI and hydroxyl radical footprinting

To study LrpA binding at its own promoter in more detail, we performed DNaseI footprinting with purified LrpA and fragment B containing the *lrpA* promoter (Fig. 1B). Addition of LrpA resulted in DNaseI protection at a region of -19 to +21 relative to the *lrpA* transcriptional start (+1, Fig. 7A). This indicated that this region contained the LrpA binding sequence. Additionally, sites hypersensitive to DNaseI cleavage were observed. Such an effect is common when DNA-binding proteins bind to their target DNA, and can be explained as protein-induced DNA bending (27). In the case of LrpA, however, apparent hypersensitivity appeared mainly at positions outside of the -19 to +21 region, suggesting that binding of LrpA affected the DNA conformation upstream and downstream of its binding site (see discussion below). DNaseI footprinting experiments were also performed with purified LrpA and fragment G containing the *gdh* promoter. As in gel-mobility shift experiments, no interaction between LrpA and *gdh* promoter DNA could be detected (not shown).

To further characterize the LrpA binding site(s) at its own promoter, hydroxyl radical footprinting was performed with the *lrpA* promoter fragment B (Fig. 1B). Using this technique information can be obtained about the interactions between LrpA and the DNA sugar-phosphate backbone. The plot profiles of the relative band intensities from these experiments revealed the presence of four regions with decreased band intensities in both the non-template and the template strand (Fig. 7B). These regions are all located between

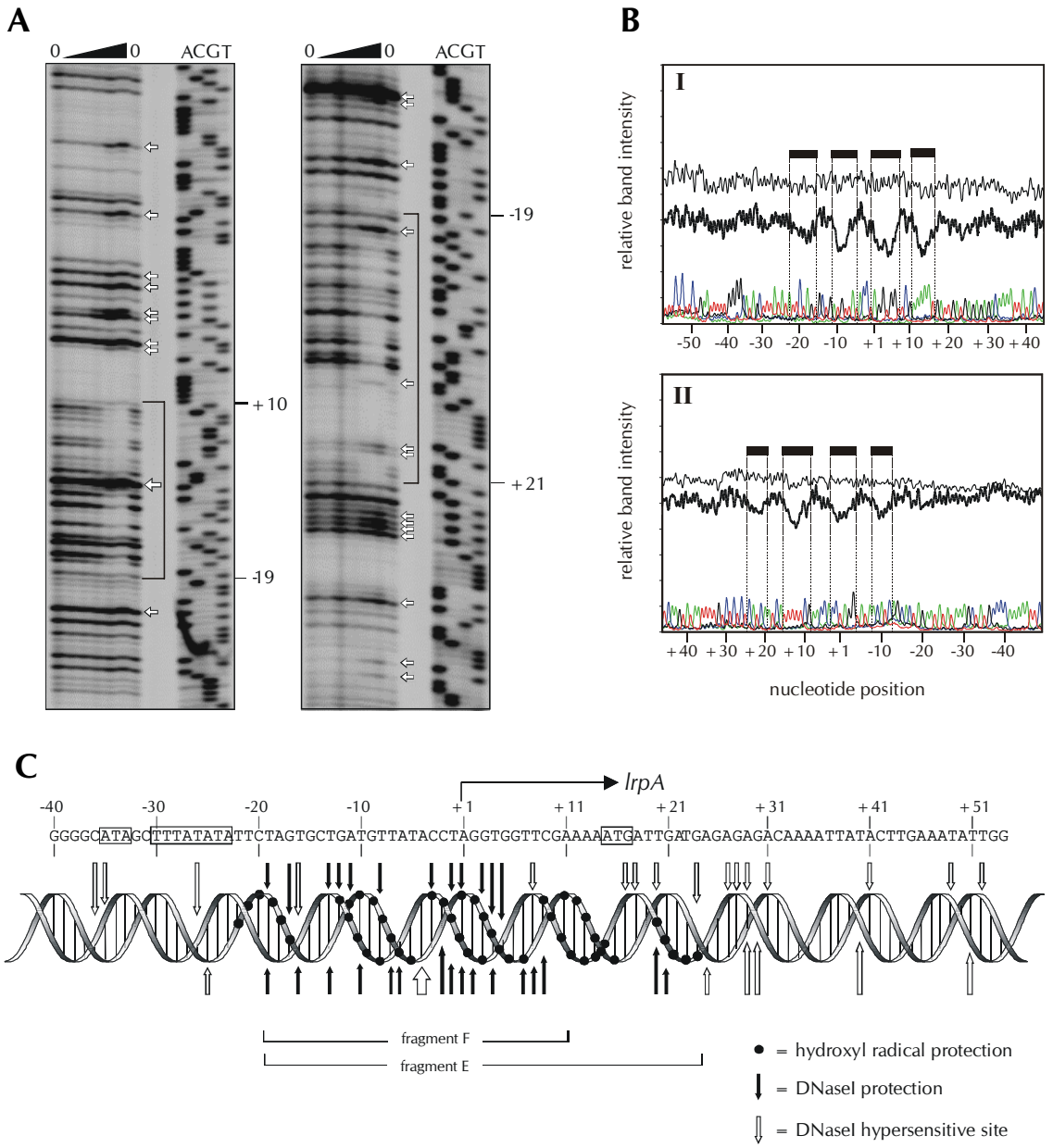



**Figure 7.** (**A**) DNaseI footprint analysis using purified LrpA and *lrpA* promoter DNA. 0, 30, 90, 270, 810, 2430, and 0 nM of purified LrpA was incubated with labeled fragment B (see Fig. 1B). After DNaseI treatment the DNA was analyzed in parallel with a sequencing reaction. Left, 5′-labelled non-template strand; right, 5′-labelled template strand. Brackets indicate DNaseI protection, and arrows refer to hypersensitive sites. (**B**) Profile plots of hydroxyl radical footprinting. Panel I, non-template strand; panel II, template strand. Nucleotide positions are given relative to the *lrpA* transcriptional start (+1). Upper lines, no LrpA added; lower lines (bold), 2.4 µM LrpA added, as calculated for monomeric protein. Profile plots at the bottom indicate sequence reactions that were run in parallel with hydroxyl radical cleavage reactions. Brackets indicate protection against hydroxyl radical cleavage in the presence of LrpA. (**C**). Schematic summary of results obtained from gel-mobility shift experiments, DNaseI footprinting and hydroxyl radical footprinting. A double helical representation of the *lrpA* promoter region is given along the sequence of the non-template strand. Filled circles, protection against hydroxyl radical cleavage; filled arrows, protection against DNaseI cleavage; open arrows, DNaseI hypersensitive sites. Sequence positions are relative to the *lrpA* transcriptional start (+1). Boxed sequences indicate the BRE-element, TATA-element, and ATG-startcodon, respectively

positions -22 and +24. A spacing of approximately 10 bp was present between disappearing bands, which indicated that interactions took place along the same face of the double helix. A schematic summary of the data obtained from gel-mobility shifts, DNaseI and hydroxyl radical footprinting is given in Fig. 7C.

## Mutational analysis of the LrpA binding sequence

Because most of the protection against DNaseI cleavage occurred within the -19 to +11 region, we tested whether LrpA was able to shift a 30-bp fragment (F, Fig. 7), containing only this -19 to +11 sequence. This fragment is only partly shifted, even at higher LrpA concentrations (Fig. 6). Although binding to this fragment was thus much weaker than to larger fragments like E or D, sequence elements specifically recognized by LrpA are present within this region. Therefore, we designed mutations in this particular region for a more detailed analysis of the sequence elements required for the LrpA-DNA interaction. The effects of these mutations were tested in gel-mobility shift experiments and in the cell-free transcription system mentioned above. In gel-mobility experiments we used mutants of fragment B (Fig. 1B), and for the cell-free transcription experiments we cloned mutants of fragment A (Fig. 1B) into pGEM-T. Although transcriptional activities or transcriptional start sites were slightly altered for some mutant promoters, it was possible to study the effect of LrpA on transcription from the mutated templates.

We used different approaches to design mutations. In the first approach we mutated four of the eight bp of the palindromic sequence ACCTAGGT present within the -19 to +11 sequence (Fig. 8A, 621). In a gel-mobility shift experiment, however, a DNA fragment containing this mutation shifted with an efficiency almost identical to that of the wild-type DNA fragment (Fig. 8B). It was impossible, however, to analyse the effect of this mutation by *in vitro* transcription, since no transcript was formed using this mutant DNA as a template. A sequence element crucial for transcription is apparently disturbed in this mutant. Most likely this element corresponds to (part of) the initiator element (INR, previously referred to as box B) (56, 57), since it is located around the transcriptional start site.

In the second approach we compared putative *lrpA* promoter sequences of *Pyrococcus horikoshii* (30) and *Pyrococcus abyssi* (http://www.genoscope.cns.fr/cgi-bin/Pab.cgi) with that of *P.furiosus*. All these *Pyrococcus* species have a similar organization of *gdh* and *lrpA* genes, and putative promoter sequences of *lrpA* share a high degree of homology. In Fig. 8A, the bases that are conserved among the three *Pyrococcus* species are indicated with a grey background. Upstream from the TATA boxes there is little conservation, whereas a higher

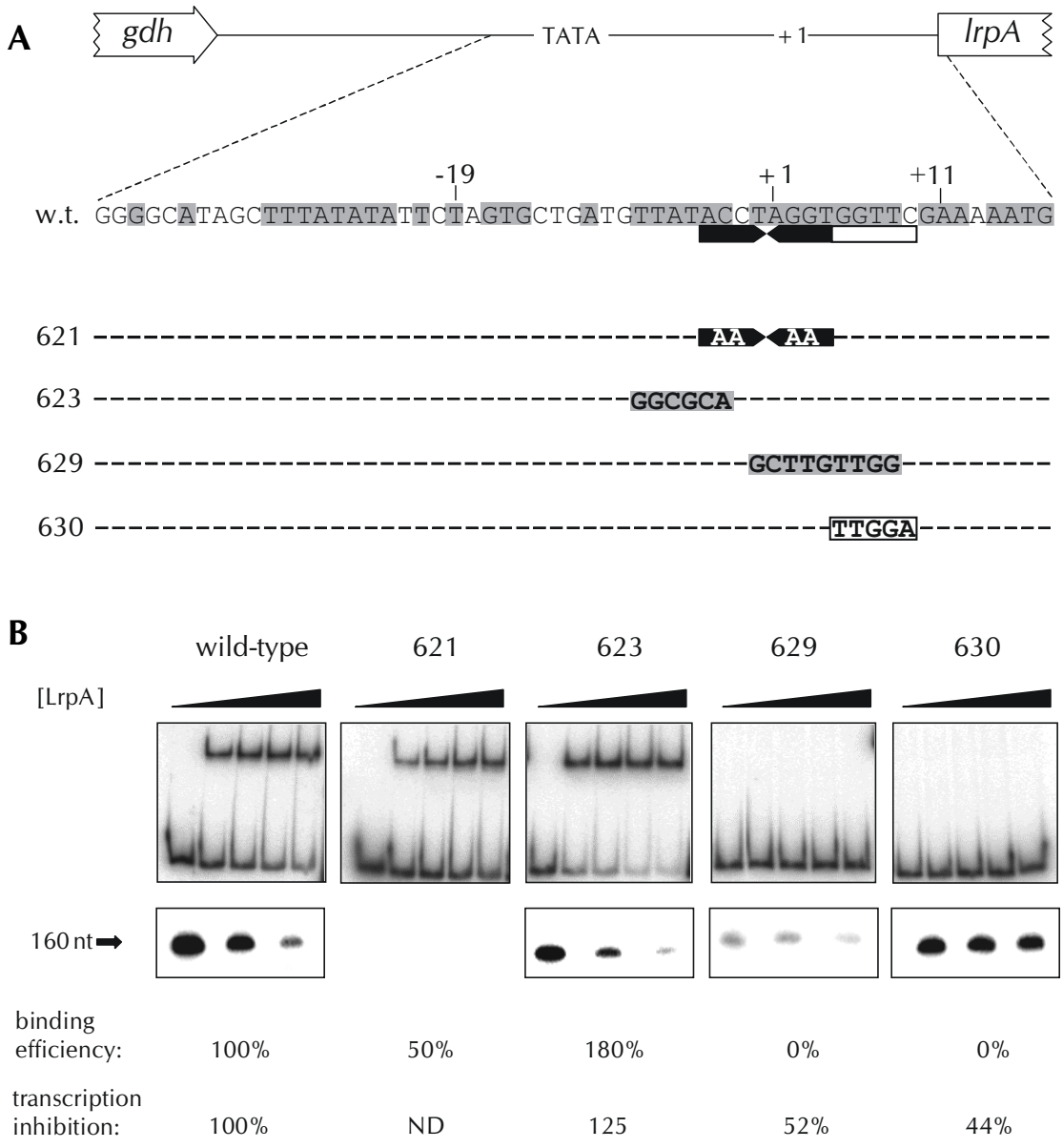



**Figure 8.** (**A**) Analysis of mutant *lrpA* promoters by mobility-shift experiments, and in *in vitro* transcription. Black filled arrows, eight-bp palindromic sequence; sequence with gray background, conserved sequence between *lrpA* promoters of *P.furiosus*, *P.horikoshii*, and *P.abyssi*. The -19 to +11 region protected from DNaseI cleavage is indicated (see Fig. 7). (**B**) LrpA-DNA binding, and *in vitro* transcription with mutant *lrpA* promoter DNA. Wild-type and mutants of fragment B (Fig. 1B) were used in gel-mobility shift experiments. LrpA concentrations were 0, 30, 90, 270, and 810 nM, as calculated for monomeric protein. 2 µg of poly(dI.dC).poly(dI.dC) was present in each reaction. To determine the percentage of shift, the bands were quantified using a phosphor imager. The plasmids pLUW613 (wild type), pLUW623, pLUW629, and pLUW630 were linearized with *Pst*I and analyzed in *in vitro* transcription in the presence of 0, 1.1, and 2.2 µM of LrpA. The arrow indicates the 160-nucleotide-run-off-transcript from the wild type. Transcription activities were calculated using a phosphor imager. Binding efficiency indicates the percentage of shift relative to the wild-type that was obtained in gel-mobility shift experiments upon addition of 30 nM LrpA. Transcription inhibition indicates the inhibition of transcription by LrpA relative to the wild-type, observed in the presence of 1.1 µM of LrpA.

degree of conservation is present around the TATA elements and the sequence downstream thereof. We mutated two conserved sequence blocks present within the -19 to +11 region (Fig. 8A, 623 and 629). When we substituted TTATAC into GGCGCA (623) we found that the affinity of LrpA for this mutant DNA was in fact higher than for the wild-type DNA sequence. In accordance, LrpA inhibited *in vitro* transcription more efficiently (125 percent, Fig. 8B). When we substituted TAGGTGGTT into GCTTGTTGG (629) there was no binding of LrpA in a gel-mobility shift experiment. This was in agreement with the affected transcription inhibition, which was only 52 percent compared to wild-type DNA (Fig. 8B).

Finally, we focused on the right half of the substituted bases in promoter mutant 629, since bases at the left half did not cause a drastic effect on LrpA binding (mutant 621). Therefore, we constructed promoter mutant 630, in which GGTTC is substituted by TTGGA (Fig. 8A). This substitution caused the same effect as in mutant 629. LrpA did not bind the mutant promoter DNA, and in agreement with this, transcription inhibition by LrpA was decreased to 44 percent compared to wild-type DNA (Fig. 8B). The observation that there was still an effect in *in vitro* transcription, although there was no detectable binding to mutant fragments 629 and 630 can be explained by the fact that nonspecific competitor DNA was only added in gel-mobility shift experiments. In the absence of competitor DNA, very weak LrpA binding still occurred (not shown).

## Chemical cross-linking

Results from gel filtration chromatography suggested that LrpA existed as a dimer, tetramer, and octamer in solution. To further characterize the configuration of LrpA as a free protein or in complex with *lrpA* promoter DNA, we performed chemical cross-linking experiments. Dimethyl suberimidate (DMSI) was used as cross-linking agent, which acts by forming covalent amide linkages between lysine residues. This results in cross-links primarily between the subunits of oligomeric proteins (12). Cross-linked LrpA was analysed by SDS-PAGE, followed by Western blotting using an antiserum raised against purified LrpA. First, cross-linking experiments were performed with LrpA as a free protein in solution. For that purpose, LrpA was incubated with DMSI, and cross-linking was analysed using SDS-PAGE. DMSI caused the appearance of three bands in addition to the LrpA monomer (I, Fig. 9) corresponding to molecular masses of an LrpA dimer (II), trimer (III) and tetramer (IV). Although the band corresponding with the dimer form (II) appears to be the predominant species, a band corresponding with the tetrameric form (IV) is present as well. While the results from gel filtration chromatography showed that LrpA also exists

as an octamer in solution, this configuration was not observed in cross-linking experiments. We tested several concentrations of LrpA (200 nM to 40 μM) under identical cross-linking conditions, but did not observe an obvious change in the intensities of the individual bands (not shown). This suggests that the degree of multimerization of LrpA is not concentration-dependent under the tested conditions.

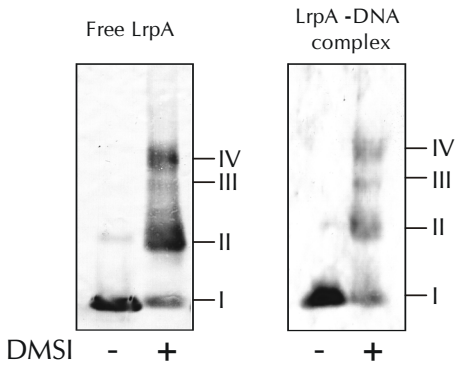



**Figure 9.** SDS-PAGE analysis of cross-linking experiments with and free LrpA (left) or LrpA in complex with DNA (right). Numbers I to IV indicate molecular masses corresponding to an LrpA monomer, dimer, trimer, and tetramer, respectively.

We also performed cross-linking experiments with LrpA bound to its target DNA. LrpA was incubated with *lrpA* promoter DNA, DMSI was added, and the samples were separated on a non-denaturing acrylamide gel. Subsequently the bands representing specific LrpA-DNA complexes were excised and analysed by SDS-PAGE. The pattern of cross-linked LrpA in complex with DNA was almost similar to that of free LrpA (Fig. 9). In both cases bands with molecular masses corresponding to the dimer, trimer and tetramer or LrpA were present in addition to the monomer. However, in cross-linked LrpA that is complexed with DNA the tetrameric form appeared to be the more dominant species. Therefore, these results suggest that at least four LrpA monomers are in complex with the *lrpA* promoter fragment.

## DISCUSSION

The present study describes the characterization of the LrpA transcriptional regulator from the hyperthermophilic archaeon *Pyrococcus furiosus*. Comparison of the amino acid sequence of LrpA reveals that it belongs to the Lrp/AsnC family of transcriptional regulators, which consists of many bacterial proteins as well as a growing number of putative archaeal proteins (Fig. 2; http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palog?AF1148). Although gel-filtration experiments showed that the purified recombinant LrpA exists as a mixture of dimer, tetramer, and octamer in solution, chemical cross-linking experiments suggest that the tetrameric form of the protein was the highest quaternary structure, both in solution and in complex with DNA (Fig. 9). In comparison, *E.coli* Lrp exists as a dimer, both in solution and in complex with a single binding site.

*Pseudomonas putida* BkdR exists as a tetramer in solution, and three tetramers are proposed to bind to its target DNA (11, 28, 40, 69).

*In vitro* analyses revealed that *P.furiosus* LrpA binds to its own promoter and represses its transcription. We tested potential effectors for their ability to alter the effect of LrpA on *lrpA* transcription, however, both *in vitro* transcription analyses and gel-mobility shift assays revealed that these compounds had no effect on the LrpA autoregulation efficiency. Negative autoregulation is a common characteristic within the Lrp/AsnC family of regulatory proteins, and in general this repression is independent of effectors (29, 34, 65).

It has been reported that the RNA polymerase of the archaeon *Methanococcus thermolithotrophicus* protects a region of -30 to +20 at a number of promoters in DNaseI footprinting experiments (59). In a similar approach, Hausner *et al.* (23) demonstrated that TBP and TFB together protect the -42 to -19 region of the *P.furiosus gdh* promoter. Assuming that these proteins bind the same region at the *lrpA* promoter, LrpA binding occurs immediately downstream of the bound TBP-TFB complex (-22 to +24, Fig. 7C). In this way, LrpA blocks at least the binding of the RNA polymerase, thereby inhibiting completion of the archaeal transcription initiation complex. Such a mechanism of repression has recently been demonstrated for the *Archaeoglobus fulgidus* metal-dependent transcriptional repressor MDR1 (1). Repression occurs when MDR1 binds to the -18 to +67 sequence of its target promoter, thereby preventing RNA polymerase recruitment, but not binding of the TBP-TFB complex. Since the 5' border of the LrpA binding sequence (-24) is comparable to that of MDR1 (-18), a similar mechanism of repression is anticipated.

In analogy with *E.coli* Lrp, it is tempting to speculate that the archaeal homologues may also act as global regulators (8), and thereby act both as transcriptional repressors and activators. Activation of transcription would probably require binding upstream of the -42 to -19 region involved in binding of the TBP-TFB complex, as has been suggested for the *Haloferax* GvpE transcriptional activator (35). In addition, a probable requirement would be a specific interaction between the bacterial-like LrpA protein and one or more components of the eukaryal-like (pre-)initiation complex. Although many genes encoding Lrp/AsnC homologues can be found within Archaea, their role as transcriptional global regulators or activators remains speculative.

Hydroxyl radical footprinting analysis confirmed that LrpA protects its promoter from position -22 to +24 (Fig. 7B). We compared the size of the protected sequence (46 bp) to hydroxyl radical footprints obtained from *E.coli* Lrp in complex with *ilvIH* and *ilvGMEDA* (48, 63). In these cases two *E.coli* Lrp dimers bind to a dual binding site of about 45 bp. In addition, it has been shown that one Lrp dimer binds to a single binding site of about 15

bp (11). It is not unlikely that the *P.furiosus* LrpA binding sequence consists of two adjacent binding sites. In gel-mobility shift experiments only one distinct LrpA-DNA complex is present. This suggests either binding of an LrpA tetramer, or the highly co-operative binding of two LrpA dimers. In both cases four LrpA monomers are in complex with DNA, as was detected using chemical cross-linking of LrpA in a specific LrpA-DNA complex. Both of the above mentioned configurations (tetramer or two dimers) would explain why a fragment of at least 43 bp is necessary for a stable LrpA-DNA complex in gel-mobility shift experiments (see Fig. 6). In contrast, for a stable interaction between an *E.coli* Lrp dimer and a single binding site, a DNA fragment of only 21 bp is necessary (11).

The 46-bp binding sequence has been analysed for sequence motifs that could serve as targets for DNA binding proteins (e.g. palindromes, inverted or direct repeats). An obvious 8-bp palindrome ACCTAGGT is present (Fig. 8A), but the possibility that this is a specific recognition element for LrpA can be ruled out because LrpA binds well to promoter mutant 621, in which the palindrome has partly been substituted (Fig 8B). Moreover, the consensus binding sequence for *E.coli* Lrp consists of at least 15 bp. Assuming that LrpA binds to two sites, it might be expected that the two sites share similar sequence elements. Overall, however, there is very little sequence homology between the two halves of the 46 bp fragment. Although a CTAG motif is present in the left part of both of these halves (position -20 and -2), the combined results from our mutational analysis suggest that the GGTTC sequence and not the CTAG sequence is specifically recognized by LrpA (Fig. 8). No similar sequence is present in the other half of the 46 bp fragment. Although the LrpA binding sequence is very well conserved between *P.furiosus*, *P.horikoshii*, and *P.abyssi*, we found that substitutions in some of these conserved sequences (e.g. promoter mutant 623) had no effect on LrpA binding. It is therefore possible that LrpA does not recognize a limited number of specific bases, but rather a relatively long sequence with a specific secondary structure. Indeed, certain protein-DNA interactions have been reported to be the result of a specific DNA structure and/or flexibility (60). In addition, it should be noted that a consensus sequence for *E.coli* Lrp is in many cases not obvious at all.

The appearance of DNaseI hypersensitive sites in the proximity of DNaseI protected regions is a very common phenomenon when DNA-binding proteins interact with their target DNA, and is generally interpreted as a result of DNA bending. In some cases this bending occurs when DNA is looped out through interaction between proteins bound to sites spaced (far) apart (27). In many other cases, binding of a protein to a single binding site causes bending of DNA (33, 49, 53). DNaseI footprinting studies with these proteins,

however, revealed only a small number of hypersensitive sites, generally located within the area protected against cleavage (4, 7, 18-20, 23, 31, 38, 47, 55, 61, 67). DNaseI footprints of *E. coli* Lrp with several target promoter fragments show protection from DNaseI over a range of 100 bp or even longer, due to multiple binding sites. Hypersensitive sites are located between the protected regions (16, 21, 41, 46, 63, 68). Similar patterns are generated by the Lrp homologues BkdR from *P.putida* (40) and PutR from *Agrobacterium tumefaciens* (29). In the case of *P. furiosus* LrpA, however, two distinct footprinting techniques indicate only a significant protection of a 46-bp fragment of the *lrpA* promoter, whereas DNaseI hypersensitive sites are extended in both directions, in total covering approximately 88 bp (Fig. 7). Hence, at least under the conditions used for the DNaseI footprinting (elevated LrpA concentrations), LrpA appears to affect the DNA conformation over a relatively long distance. In the case of *E.coli* Lrp and *A.tumefaciens* PutR it has been suggested that DNA is somehow wrapped around the regulators (64, 68). Analysis of PutR-DNA complexes by atomic force microscopy supports the idea that PutR condenses DNA into globular nucleoprotein complexes (29). Although a similar looping or wrapping of DNA around *P. furiosus* LrpA is possible, we can not rule out the possibility that the observed hypersensitivity pattern is a result of some non-specific phenomenon in the *in vitro* analysis.

In conclusion, we have described for the first time the actual involvement of an archaeal Lrp homologue in transcription modulation by *in vitro* analyses. The *P. furiosus* LrpA interacts specifically with the *lrpA* promoter in the proximity of the transcriptional start site. Hence, the observed transcription inhibition is most likely a consequence of preventing RNA polymerase recruitment, similar to that reported for the *Archaeoglobus fulgidus* MDR1 (1). In addition, LrpA binds to *lrpA* promoter fragments in a single configuration, most likely as a tetramer. Alternatively, such a configuration may also be referred to as a dimer of dimers, but we do not have any indication of cooperativity, as has been reported for *E. coli* Lrp (Calvo *et al.* 1994; Newman *et al.,* 1995). Gel retardation analysis revealed that the DNA-binding efficiency of LrpA is reduced significantly when the DNA fragments were reduced in size below 46-bp. The interaction with this fragment was confirmed in two distinct footprinting experiments. The actual binding site could at least in part be identified, but as in many bacterial Lrp-target promoters, no obvious palindromic motif(s) appear to be involved. Recent progress with crystallization of *P. furiosus* LrpA (54) may be very important to confirm molecular details of the archaeal and bacterial Lrp homologues, and as such LrpA may be a model for further understanding structure-function relations of this widely distributed class of transcription regulators.

## ACKNOWLEDGEMENTS

This research was supported by grant 700-35-101 of the Council for Chemical Sciences (CW) of the Netherlands Organization for Scientific Research (NWO), by grants of the Deutsche Forschungsgemeinschaft, and the Fonds der Chemischen Industrie to M. Thomm.

## EXPERIMENTAL PROCEDURES

### DNA sequence analysis

Identification of LrpA homologues was done using the Advanced BLAST program at NCBI (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/). Alignments were made using the program ClustalX. Motif searches were performed using the PROSITE Pattern and Profile Searches program at the ExPaSy Molecular Biology Server (http://expasy.hcuge.ch/), and the program HELIX-TURN-HELIX (http://pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hth.html). Inverted repeats were identified using the GeneQuest program that is part of the DNA Star package.

### Plasmid and strain construction

The gene encoding *lrpA* was PCR amplified using primers BG240 and BG241 (see Table 1, Italic and underlined sequences indicate the restriction sites *Bsp*HI and *Bam*HI, respectively). The resulting 444-bp PCR fragment was cloned into pGEM-T (Promega, Corp.), resulting in pLUW600, and the sequence of the insert was verified by DNA sequencing. Subsequently, pLUW600 was digested with *Bsp*HI and *Bam*HI and the resulting 428-bp fragment was cloned into the T7 expression vectors pGEF+ (50), pET9d and pET24d (58) (Novagen, Inc.), resulting in the constructs pLUW601, pLUW604 and pLUW605, respectively. These constructs were transformed into *E.coli* BL21(DE3), BL21(DE3) (pLysS) and BL21(DE3) (pLysE) (Novagen, Inc.) and tested for expression (not shown). The optimal result was obtained with *E.coli* BL21(DE3) in combination with the pET9d-derivative pLUW604. This combination was used for further expression experiments. pLUW613 was made by cloning fragment A into pGEM-T (Promega, Corp.). Mutations in fragment A were introduced using *Pfu* polymerase in the PCR-based overlap extension method (25). For each mutation a sense/antisense primerpair was designed. BG730 and BG731 introduced mutation 621; BG759 and BG760 introduced mutation 623;

| Name | Sequence (5' → 3') |
|---|---|
| BG109 | TTTACAGAACCGTCATCCATTTC |
| BG240 | GCGCG*TCATGA*TTGATGAGAGAGACAAAATTATAC |
| BG241 | CGCGC*GGATCC*TTACTTAAGTTTTTCAAGGATTATAG |
| BG289 | CAGAACATAACTGGATACTACTGGA |
| BG290 | CAAGAGCCTTGACTCTCTTCCTC |
| BG367 | GGATGGGTCAAGCACTGATTCC |
| BG427 | CTCTAGAATGTTCAACACTATGGCTC |
| BG430 | GGGGCATAGCTTTATATATTCTAGTGCTGAT |
| BG431 | ATCAGCACTAGAATATATAAAGCTATGCCCC |
| BG498 | CATCAATCATTTTTCGAACCACCTAGGTATAAC |
| BG615 | TCGAACCACCTAGGTATAACAT |
| BG638 | TAGTGCTGATGTTATACCTA |
| BG730 | GCTGATGTTATAAATAAATGGTTCGAAAAA |
| BG731 | TTTTTCGAACCATTTATTTATAACATCAGC |
| BG759 | TAGTGCTGATGGGCGCACTAGGTGGTTCGA |
| BG760 | TCGAACCACCTAGTGCGCCCATCAGCACTA |
| BG792 | GTTATACCGCTTGTTGGCGAAAAATGATTGATGAG |
| BG793 | TTTTTCGCCAACAAGCGGTATAACATCAGCACTAG |
| BG794 | ATACCTAGGTTTGGAGAAAAATGATTGATGAGAG |
| BG795 | AATCATTTTTCTCCAAACCTAGGTATAACATC |

**Table I.** Oligonucleotides used in this study

BG792 and BG793 introduced mutation 629; BG794 and BG795 introduced mutation 630 (see Table 1). BG289 and BG290 were used as flanking primers for the PCR of fragment A (see below). All mutant fragments A were cloned into pGEM-T (Promega, Corp.) and sequenced, resulting in pLUW621, pLUW623, pLUW629, and pLUW630.

## Overproduction of LrpA

The *P.furiosus* LrpA protein was produced in 2-liter Erlenmeyer flasks, containing 1 liter LB medium with 50 μg/ml kanamycin. The culture was inoculated with *E.coli* BL21(DE3) containing pLUW604. Cells were grown in a rotary shaker at 37 °C until an $OD_{600}$ of 0.5 was reached, and 0.4 mM IPTG was added to induce expression. After O/N incubation the cells were harvested, washed in 125 mM citrate buffer pH 5.0 and resuspended in 90 ml of the same buffer. Cells were lysed by a triple passage through a French pressure cell at 1000 psi. After lysis $MgCl_2$ and DNaseI were added to final concentrations of 10 mM and 10 μg/ml, respectively. The sample was left at room temperature for 15 min. Subsequently, the cell-free extract was incubated at 80°C for 30 min and centrifuged at 20,000 rpm for 30 min. The remaining soluble fraction was loaded on a 60-ml cation exchange column (S-sepharose, Pharmacia) that had been equilibrated with 125 mM citrate buffer pH 5.0. The column was eluted with the same buffer, using a flow rate of 3 ml/min. and a linear gradient of NaCl from 0 to 1 M. Fractions containing LrpA, as determined by SDS-PAGE, were pooled and concentrated by centrifugation in Centricon units (10 kDa cut-off), until a volume of 0.5 ml was reached. A 200 μl sample was loaded on a gel filtration column (Superdex 200, Pharmacia) with 20 mM Tris pH 8.0 and 100 mM NaCl with a flowrate of 0.5 ml/min. The elution pattern from this gel filtration showed 3 peaks, corresponding to molecular masses of approximately 30 kDa , 60 kDa, and 120 kDa, respectively. Approximately 14 mg of purified LrpA was obtained from one litre of culture.

## Gel-mobility shift experiments

DNA probes used for gel-mobility shift experiments were generated using PCR. The following primers were used: BG289 and BG290 for fragment A; BG367 and BG290 for fragment B; BG289 and BG431 for fragment C; BG430 and BG290 for fragment D; BG638 and BG498 for fragment E; BG638 and BG615 for fragment F; BG427 and BG109 for fragment G (see Table 1). PCR reactions consisted of a 5 min. denaturation step at 95ºC, 30 cycles consisting of 95ºC, 45ºC and 72ºC, with 30 seconds for each step, followed by a 7 min. final extension step at 72ºC. PCR products were end-labelled using T4 kinase and radioactive $\gamma^{32}$P-ATP. Binding reactions were performed in a total volume of 20 µl, containing 40 mM HEPES-NaOH pH 7.3, 200 mM KCl, 2.5 mM $MgCl_2$, 2 mM DTT, 1 mM $CaCl_2$, 100 mM EDTA, 10% glycerol, and varying concentrations of purified LrpA. Standard reactions contained 2 µg of poly(dI.dC).poly(dI.dC) as nonspecific competitor DNA, but this was omitted from reactions with smaller fragments (fragment E and F) and during determination of the dissociation constant ($K_d$) for the LrpA-DNA complex. Each reaction contained 1 to 10 ng of $\gamma^{32}$P-ATP end-labelled DNA. Reactions were incubated at room temperature for at least 10 min. and separated on a non-denaturing 8% acrylamide gel, buffered in 1x Tris Borate EDTA buffer (51). In the case of fragment E a 20% gel was used. Gels were dried, exposed to phosphor screens and analysed. Quantification was done using Imagequant software (Molecular Dynamics, Inc.).

## DNaseI and hydroxyl radical footprinting

DNaseI and hydroxyl radical footprinting was performed using non-radioactive probes containing the IRD800 label, in combination with a Li-Cor sequencer (Li-Cor, Inc.). For this purpose DNA probes were prepared as follows; fragments B and G (Fig. 1B) were cloned into pGEM-T (Promega, Corp.), and clones were selected containing the insert in both orientations. These constructs were used as a template in PCR reactions with a IRD800-labelled T7 primer (MWG-Biotech, GmbH) in combination with BG290, BG367, BG427 or BG109. These PCR reactions produced fragments B or G carrying the IRD800 label on a 68 bp extension originating from pGEM-T at the 5' end of either the non-template or the template strand. 10 ng of this DNA was used per reaction. Binding reactions for DNaseI and hydroxyl radical footprinting were identical to the binding reaction conditions in gel-mobility shift experiments (see above) except that glycerol and poly(dI.dC).poly(dI.dC) were omitted. DNaseI cleavage was done by adding 20 µl of a solution containing 5 mM $CaCl_2$, 10 mM $MgCl_2$ and 2 mU of DNaseI. After 1 min. the

DNaseI reaction was stopped by the addition of 20 µl 4 M $NH_4Ac$ and 30 mM EDTA. The DNA was extracted with 60 µl phenol, precipitated with 96% ethanol in the presence of 20 µg glycogen and washed with 70% ethanol. The pellet was dissolved in 1 µl of formamide loading buffer, heated at 95°C for 5 min. and chilled on ice. Subsequently, 0.8 µ l was analysed on a Li-Cor 4000 sequencer (Li-Cor, Inc.) using a 5.5% *KB*$^{Plus}$ 41 cm denaturing sequence gel (Li-Cor, Inc) with 0.2 mm spacers and settings 2000 V, 25 mA, 50 W and 45°C.

Hydroxyl radicals were generated by adding 3 µl of 40 mM Na-Ascorbate, 3 µl of 1.2% $H_2O_2$, and 3 µl of 4 mM $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$, 8 mM EDTA. After 2 min. the reaction was stopped by the addition of 26 µl of 0.1 M Thiourea, 20 mM EDTA. DNA was extracted with phenol and precipitated as described above, and analysed on a Li-Cor 4000L sequencer (Li-Cor, Inc.), using a 66 cm denaturing sequencing gel with 0.25 mm spacers and settings 2250 V, 30.6 mA, 68 W and 45°C. Images of the footprints were analysed using the program Scion Image for Windows (http://rsb.info.nih.gov/nih-image/).

### *In vitro* transcription

Transcription reactions were performed essentially as described previously (24), except that 300mM KCl was used instead of 250 mM. A standard reaction mixture (50µl) contained 1µg linearized template DNA (pLUW479 (24), pLUW613, pLUW621, pLUW623, pLUW629, pLUW629, or piC31/2 (22)), 250 ng recombinant TBP, 280 ng recombinant TFB, 135 ng native RNA polymerase, and varying concentrations of LrpA. This reaction mixture was incubated for 30 min. at 70°C. RNA purification and electrophoresis was performed as described previously (17).

### Primer extension

For analysis of the *in vitro* transcriptional start site, cell-free transcription reactions were performed as described above but with unlabelled precursors. In a control reaction, nucleotides were omitted from cell-free transcription reactions. The end-labelled DNA primer 5'-GTATAATTTTGTCTCTCTCATCA-3' was used, complementary to nucleotides +20 to +42 relative to the transcriptional start of *lrpA*. The primer extension assay was performed as previously described (17, 24).

For determination of the *in vivo* transcriptional start site, primer extension was performed with total RNA of *P.fusiosus*, which was isolated as described previously (66).

## Chemical cross-linking and Western blot analysis

Chemical cross-linking was performed as described by Davies *et al.* (12), with the following modifications. For cross-linking experiments with free LrpA, different concentrations of LrpA were diluted in cross-linking buffer (80 mM triethanolamine-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.1 mM DTT) and the final volume was adjusted to 16 µl. Dimethylsuberimidate (DMSI, 25 mg/ml freshly made in cross-linking buffer) was added to a final concentration of 5 mg/ml so that the final volume was 20 µl. After a one-hour incubation at room temperature SDS-PAGE loading buffer (51) was added, and 100 ng of each sample was separated on a 10% Tricine SDS-PAGE gel (52). The separated proteins were transferred to a nitrocellulose membrane by electroblotting in 10 mM CAPS pH 11.0 and 10% methanol, and detected immunologically using a polyclonal antiserum raised against purified LrpA (51). For cross-linking experiments with LrpA-DNA complexes, about 2 µg of purified LrpA was incubated with 500 ng of DNA (fragment B, see Fig. 1B) in cross-linking buffer in a final volume of 64 µl. DMSI was added as described above so that the final volume was 80 µl. The samples were loaded on a non-denaturing 5% acrylamide gel, buffered in 1x Tris Borate EDTA buffer (51). The gel was stained with ethidium bromide and bands representing specific DNA-LrpA complexes were excised, crushed, and SDS-PAGE loading buffer was added. The recovered samples were heated for 10 min. at 100°C, and loaded on a 10% Tricine SDS-PAGE gel and analysed as described above.

## REFERENCES


1. Bell, S.D., S.S. Cairns, R.L. Rob son, and S.P. Jackson 1999. Transcriptional regulation of an archaeal operon *in vivo* and *in vitro*. Mol Cell. 4:971-982.
2. Bell, S.D., and S.P. Jackson 1998. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends in M icrobiology. 6:222-228.
3. Bell, S.D., P.L. Kosa, P.B. Sigler, and S.P. Jackson 1999. Orientation of the transcription preinitiation complex in archaea. Proc Natl Acad Sci. 96:13662 -13667.
4. Bernues, J., Carrera, P., Azorin, F. 1996. TBP binds the transcr iptionally inactive TA5 sequence but the resulting complex is not efficiently recognised by TFIIB and TFIIA. Nucleic Acids Res. 24:2950 -2958.
5. Calvo, J.M., and R.G. Matthews 1994. The leucine -responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiol Rev. 58:466-490.
6. Carey, J. 1991. Gel Retardation. Meth Enzymology. 208:103 -117.
7. Chapon, C., Kolb, A. 1983. Action of CAP on the *malT* promoter *in vitro*. J Bacteriol. 156:1135-1143.
8. Charlier, D., Roovers, M., Thia-Toong, T.L., Durbecq, V., Glansdorff, N. 1997. Cloning and identification of the *Sulfolobus solfataricus lrp* gene encoding an archaeal homologue of the eubacterial leucine-responsive global transcriptional regulator Lrp. Gene. 201:63 -68.
9. Cohen-Kupiec, R., Blank, C., Leigh, J.A. 1997. Transcriptional regulation in Archaea: *in vivo* demonstration of a repressor binding site in a methanogen. Proc Natl Acad Sci. 94:1316 -1320.
10. Cohen-Kupiec, R., Marx, C.J., Leigh, J.A. 1999. Function and regulation of *glnA* in the methanogenic archaeon *Methanococcus maripaludis* J Bacteriol. 181:256-261.
11. Cui, Y., Midkiff, M.A., Wang, Q., J.M. Calvo. 1996. The leucine -responsive regulatory protein (Lrp) from *Escherichia coli*, stoichiometry and minimal requirements for binding to DNA. J Biol Chem. 271:6611-6617.
12. Davies, G.E., and G.R. Stark 1970. Use of dimethyl suberimidate, a cross -linking reagent, in studying the subunit structure of oligomeric proteins. Proc Natl Acad Sci. 66:651 -656.
13. de Wind, N., de Jong, M., Meij er, M., Stuitje, A.R. 1985. Site-directed mutagenesis of the *Escherichia coli* chromosome near *oriC*: identification and characterization of *asnC*, a regulatory element in *E. coli* asparagine metabolism Nucleic Acids Res. 13:8797 -8811.
14. Dodd, I.B., and J.B. Egan 1990. Improved detection of helix -turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. 18:5019 -26.
15. Eggen, R.I., A.C. Geerling, K.G. Waldkotter, G. Antranikian, W.M. de Vos. 1993. The glutamate dehydrogenase-encoding gene of the hyperthermophilic archaeon *Pyrococcus furiosus*: sequence, transcription and analysis of the deduced amino acid sequence Gene. 132:143 -148.
16. Ferrario, M., B.R. Ernsting, D.W. Borst, D.E. Wiese 2 nd, R.M. Blumenthal, R.G. Matthews. 1995. The leucine-responsive regulatory protein of *Escherichia coli* negatively regulates transcription of *ompC* and *micF* and positively regulates translation of *ompF*. J Bacteriol. 177:103-113.
17. Frey, G., M. Thomm, B. Brüdigam, H.P. Gohl, and W. Hausner 1990. An archaebacterial cell-free transcription system. The expression of tRNA genes from *Methanococcus vannielii* is mediated by a transcription factor. Nucleic Acids Res. 18:1361 -1367.

18. Galasinski, S.K., T.N. Lively, A. Grebe De Barron, J.A. Goodrich. 2000. Acetyl coenzyme A stimulates RNA polymerase II transcription and promoter binding by transcription factor IID in the absence of histones. Mol Cell Biol. 20:1923-1930.
19. Gaston, K., A. Bell, S. Busby, M. Fried. 1992. A comparison of the DNA bending activities of the DNA binding proteins CRP and TFIID. Nucleic Acids Res. 20:3391-3396.
20. Gaston, K., A. Bell, A. Kolb, H. Buc, S. Busby. 1990. Stringent spacing requirements for transcription activation by CRP. Cell 62:733-743.
21. Gazeau, M., F. Delort, M. Fromant, P. Dessen, S. Blanquet, P. Plateau. 1994. Structure-function relationship of the Lrp-binding region upstream of *lysU* in *Escherichia coli* J Mol Biol. 241:378-389.
22. Hausner, W., G. Frey, and M. Thomm 1991. Control regions of an archaeal gene. A TATA box and an initiator element promote cell-free transcription of the tRNA(Val) gene of *Methanococcus vannieli* J Mol Biol. 222:495-508.
23. Hausner, W., J. Wettach, C. Hethke, M. Thomm. 1996. Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. J Biol Chem. 271:30144-30148.
24. Hethke, C., A.C. Geerling, W. Hausner, W.M. de Vos, and M. Thomm 1996. A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus* Nucleic Acids Res. 24:2369-2376.
25. Higuchi, R., B. Krummel, R.K. Saiki. 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.
26. Hochheimer, A., R. Hedderich, R.K. Thauer. 1999. The DNA binding protein Tfx from *Methanobacterium thermoautotrophicum*: structure, DNA binding properties and transcriptional regulation Mol Microbiology. 32:641-650.
27. Hochschild, A. 1991. Detecting cooperative protein-DNA interactions and DNA loop formation by footprinting. Meth Enzymol. 208:343-361.
28. Huang, N., K.T. Madhusudhan, J.R. Sokatch. 1996. Stoichiometry of BkdR to substrate DNA in *Pseudomonas putida*. Biochem Biophy. Res Commun. 223:315-319.
29. Jafri, S., S. Evoy, K. Cho, H.G. Craighead, S.C. Winans. 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.
30. Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyperthermophilic archaebacterium, *Pyrococcus horikoshii* OT3 DNA Res. 5:55-76 and 147-155.
31. Kelley, W.L., D. Bastia. 1991. Conformational changes induced by integration host factor at origin gamma of R6K and copy number control J Biol Chem. 266:15924-15937.
32. Keuntje, B., B.Masepohl, W. Klipp. 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.
33. Kim, Y., J.H. Geiger, S. Hahn, P.B. Sigler. 1993. Crystal structure of a yeast TBP/TATA-box complex. Nature. 365:512-520.

34. Kölling, R., and H. Lother 1985. AsnC: an autogenously regulated activator of asparagine synthetase A transcription in *Escherichia coli*. J Bacteriol. 164:310-5.
35. Krüger, K., T. Hermann, V. Armbruster, F. Pfeifer. 1998. The transcriptional activator GvpE for the halobacterial gas vesicle genes resembles a basic region leucine-zipper regulatory protein. J Mol Biol. 279:761-771.
36. Kyrpides, N. C., C.A. Ouzounis. 1995. The eubacterial transcriptional activator Lrp is present in the archaeon *Pyrococcus furiosus*. Trends Biochem Sci. 20:140-141.
37. Kyrpides, N. C., C.A. Ouzounis. 1999. Transcription in archaea Proc Natl Acad Sci. 96:8545-8550.
38. Liljelund, P., C.J. Ingles, J. Greenblatt. 1993. Altered promoter binding of the TATA box-binding factor induced by the transcriptional activation domain of VP16 and suppressed by TFIIA. Mol Gen Genet. 241:694-699.
39. Littlefield, O., Y. Korkhin, and P. B. Sigler 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. Proc Natl Acad Sci. 96:13668-13673.
40. Madhusudhan, K.T., N. Huang, and J.R. Sokatch 1995. Characterization of BkdR-DNA binding in the expression of the bkd operon of *Pseudomonas putida*. J Bacteriol. 177:636-641.
41. Mathew, E., J. Zhi, and M. Freundlich 1996. Lrp is a direct repressor of the *dad* operon in *Escherichia coli*. J Bacteriol. 178:7234-7240.
42. Morgan, R.M., T.D. Pihl, J. Nolling, J.N. Reeve. 1997. Hydrogen regulation of growth, growth yields, and methane gene transcription in *Methanobacterium thermoautotrophicum* deltaH. J Bacteriol. 179:889-898.
43. Napoli, A., J. van der Oost , C.W. Sensen, R.L. Charlebois, M. Rossi, and M. Ciaramella 1999. An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. J Bacteriol. 181:1474-1480.
44. Newman, E.B., Lin, R. 1995. Leucine-responsive regulatory protein: a global regulator of gene expression in *E. coli*. Ann Rev Microbiol. 49:747-775.
45. Nölling, J., J.N. Reeve, 1997. Growth- and substrate-dependent transcription of the formate dehydrogenase (*fdhCAB*) operon in *Methanobacterium thermoformicicum* Z-245. J Bacteriol. 179:899-908.
46. Nou, X., B.A. Braaten, L. Kaltenbach, L., D.A. Low. 1995. Differential binding of Lrp to two sets of *pap* DNA binding sites mediated by PapI regulates Pap phase variation in *Escherichia coli*. EMBO J. 14:5785-5797.
47. Plumbridge, J., A. Kolb. 1991. CAP and Nag repressor binding to the regulatory regions of the *nagE-B* and *manX* genes of *Escherichia coli*. J Mol Biol. 217:661-679.
48. Rhee, K.Y., B.S. Parekh, G.W. Hatfield. 1996. Leucine-responsive regulatory protein-DNA interactions in the leader region of the *ilvGMEDA* operon of *Escherichia coli*. J Biol Chem. 271:26499-26507.
49. Rice, P.A., S. Yang, K. Mizuuchi, H.A. Nash. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell. 87:1295-1306.
50. Rink, R., M. Fennema, M. Smids, U. Dehmel, D.B. Janssen. 1997. Primary structure and catalytic mechanism of the epoxide hydrolase from *Agrobacterium radiobacter* AD1. J Biol Chem. 272:14650-14657.
51. Sambrook, J., E.F. Fritsch, T. Maniatis. 1989. Molecular cloning, a laboratory manual, 2nd edition. Cold Spring Harbor Laboratory Press.
52. Schagger, H., G. Jagow. 1987. Tricine-Sodium Dodecyl Sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Analytical Biochemistry. 166:368-379.

53. Schultz, S. C., G.C. Shields, and T.A. Steitz 1991. Crystal structure of a CAP -DNA complex: the DNA is bent by 90 degrees. Science. 253:1001 -1007.
54. Sedelnikova, S.E., S.H. Smits, P.M. Leonard, A.B. Brinkman, J. van der Oost, J.B. Rafferty, and D.W. Rice 2001. Crystallization and quaternary structure analysis of an Lrp -like regulatory protein from the hyperthermophile *Pyrococcus furiosus*. Acta Crystallogr D Biol Crystallogr. 57:886 -8.
55. Sirko, A., E. Zehelein, M. Freundlich, G. Sawers. 1993. Integration host factor is required for anaerobic pyruvate induction of *pfl* operon expression in *Escherichia coli* J Bacteriol. 175:5769 -5777.
56. Soppa, J. 1999. Normalized nucleotide frequenc ies allow the definition of archaeal promoter elements for different archaeal groups and reveal base -specific TFB contacts upstream of the TATA box. Mol Microbiology. 31:1589 -1592.
57. Soppa, J. 1999. Transcription initiation in Archaea: facts, factors and future aspects. Mol Microbiology. 31:1295-1305.
58. Studier, F.W., A.H. Rosenberg, J.J. Dunn, J.W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Meth Enzymol. 185:60 -89.
59. Thomm, M. 1996. Archaeal transcription factors and their role in transcription initiation. FEMS Microbiol Rev . 18:159 -171.
60. Travers, A.A. 1989. DNA conformation and protein binding Ann Rev Biochem. 58:427 -452.
61. Tsai, M.M., Y.H. Fu, R.C. Deonier. 1990. Intrinsic bends and integration host factor binding at F plasmid *oriT*. J Bacteriol. 172:4603 -4609.
62. Voorhorst, W.G., Y. Gueguen, A.C. Geerling, G. Schut, I. Dahlke, M. Thomm, J. van der Oost, W.M. de Vos. 1999. Transcriptional regulation in the hyperthermophilic archaeon *Pyrococcus furiosus*: coordinated expression of divergently oriented genes in response to beta -linked glucose polymers. J Bacteriol. 181:3777 -3783.
63. Wang, Q., and J.M. Calvo 1993. Lrp, a global regulatory protein of *Escherichia coli*, binds cooperatively to multiple sites and ac tivates transcription of *ilvIH*. J Mol Biol. 229:306 -318.
64. Wang, Q., and J.M. Calvo 1993. Lrp, a major regulatory protein in *Escherichia c*oli, bends DNA and can organize the assembly of a higher -order nucleoprotein structure. EMBO J. 12:2495 -2501.
65. Wang, Q., J. Wu, D. Friedberg, J. Plakto, J.M. Calvo. 1994. Regulation of the *Escherichia coli lrp* gene. J Bacteriol. 176:1831 -1839.
66. Ward, D.E., S.W. Kengen, J. van Der Oost, and W.M. de Vos 2000. Purification and characterization of the alanine aminotra nsferase from the hyperthermophilic Archaeon *pyrococcus furiosus* and its role in alanine production J Bacteriol. 182:2559 -66.
67. Webster, C., K. Gaston, S. Busby. 1988. Transcription from the *Escherichia coli melR* promoter is dependent on the cyclic AMP r eceptor protein. Gene. 68:297 -305.
68. Wiese, D.E. 2nd., B.R. Ernsting, R.M. Blumenthal, and R.G. Matthews. 1997. A nucleoprotein activation complex between the leucine -responsive regulatory protein and DNA upstream of the *gltBDF* operon in *Escherichia coli*. J Mol Biol. 270:152 -168.
69. Willins, D.A., C.W. Ryan, J.V. Platko, J.M. Calvo. 1991. Characterization of Lrp, and *Escherichia coli* regulatory protein that mediates a global response to leucine. J Biol Chem. 266:10768 -10774.

# Chapter 5

## Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*

Philip M. Leonard
Sander H.J. Smits
Svetlana E. Sedelnikova
Arie B. Brinkman
Willem M. de Vos
John van der Oost
David W. Rice
John B. Rafferty

**The LrpA protein from the hyperthermophilic archaeon *Pyrococcus furiosus* belongs to the Lrp/AsnC family of transcriptional regulatory proteins, of which the *Escherichia coli* leucine-responsive regulatory protein is the archetype. Its crystal structure has been determined at 2.9 Å resolution and is the first for a member of the Lrp/AsnC family, as well as one of the first for a transcriptional regulator from a hyperthermophile. The structure consists of an N-terminal domain containing a helix-turn-helix (HTH) DNA-binding motif, and a C-terminal domain of mixed α/β character reminiscent of a number of RNA- and DNA-binding domains. *Pyrococcus furiosus* LrpA forms a homodimer mainly through interactions between the antiparallel β-sheets of the C-terminal domain, and further interactions lead to octamer formation. The LrpA structure suggests how the protein might bind and possibly distort its DNA substrate through use of its HTH motifs and control gene expression. A possible location for an effector-binding site is proposed by using sequence comparisons with other members of the family coupled to mutational analysis.**

*EMBO journal 2001, 20(5): 990-997*
*(P.M. Leonard and S.H.J. Smits contributed equally to this work)*

## INTRODUCTION

Proteins from the Lrp/AsnC family, which act as global or specific regulators of transcription, have been isolated from many prokaryotes, including both bacteria and archaea (3). The most extensively studied example of this family of proteins is the leucine-responsive regulatory protein (Lrp) from *Escherichia coli* (5, 22). Lrp is a global regulator that acts to control gene expression. The Lrp regulon consists of 75 transcriptional units, which are either activated or repressed by Lrp, often in response to the presence or absence of the effector leucine, which *E.coli* Lrp is believed to bind (5). The proteins they encode are mainly involved in transport, degradation or biosynthesis of amino acids. Lrp has been shown to exhibit negative, leucine-independent autoregulation, by binding upstream of its own promoter (-80 to -32 relative to the transcription start site as determined by DNaseI footprinting; (35)). This large footprint region is believed to encompass a number of distinct Lrp binding sites. Lrp interacts with DNA as a homodimer, recognizing a 15-bp imperfect inverted repeat sometimes found in multiple copies and bound in a cooperative manner (5, 7, 22, 34). *Escherichia coli* Lrp shows notable sequence similarity to the *E.coli* AsnC protein (25% identity) and, as a result, an evolutionary family relationship between the two proteins has been proposed (36). AsnC is responsible for the asparagine-dependent regulation of the *asnA* gene, the structural gene for asparagine synthetase A, and for its own autoregulation (8, 15). No structures of proteins belonging to this family have been previously reported.

In *Pyrococcus furiosus* a putative Lrp, LrpA, the product of the *lrpA* gene, which exhibits 28% sequence identity to *E.coli* Lrp, has been isolated (3). Gel filtration experiments with concentrated protein samples suggest that LrpA forms a mixture of dimeric, tetrameric and octameric species at neutral pH, and an octamer below pH 6.0 (31). *Pyrococcus furiosus* LrpA has also been shown to exhibit negative autoregulation and binds to the *lrpA* promoter at a single site (-22 to +24 relative to the transcription start site; (3)), as determined by DNaseI and hydroxyl radical footprinting. Although apparently quite a large binding site, attempts at trimming down its size from 46 to 30 bp, encompassing the most strongly protected region, result in a substantial decrease in binding by LrpA (3). Thus, multiple copies of LrpA may bind and possibly distort the *lrpA* promoter region as suggested for its *E.coli* homologue (34). The negative autoregulation exhibited by LrpA appears to be independent of effectors, and there is no evidence for binding of leucine or any other amino acid by LrpA (3).

A helix-turn-helix (HTH) motif is responsible for the specific DNA interaction of many transcriptional regulators, such as the *E.coli* catabolite activator protein (20) and the tryptophan repressor (30). Sequence alignments of proteins belonging to the Lrp/AsnC family as well as detailed mutagenesis studies of *E.coli* Lrp have suggested that these proteins also utilize an archetypal HTH motif to interact with (27). In *P.furiosus* LrpA this motif has been predicted to be located between residues 21 and 40 (3).

This paper reports the structure determination to 2.9 Å resolution of *P.furiosus* LrpA, a description of its overall fold, its structural similarity to other proteins and the possible mode of interaction between LrpA and DNA. The LrpA structure is the first example of a member of the Lrp/AsnC family, and one of the first transcriptional regulators from archaeal or hyperthermophilic origin. The LrpA structure provides insights into the possible location of an effector-binding site and how this widely conserved prokaryotic transcriptional regulator controls gene expression.

## *RESULTS*

### Overall structure

The LrpA subunit has overall dimensions 60 x 30 x 45 Å and comprises two domains (Fig. 1A). The N-terminal domain is formed from three α-helices (αA to αC). The C-terminal domain is formed from a four-stranded anti-parallel β-sheet (β2-β5) flanked on one face by two α-helices (αD to αE) and a short C-terminal β-strand (β6). The subunit has the connectivity β2-αD-β3-β4-E-αβ5-β6 (Fig. 1B). Two $3_{10}$ helical turns are present, one between β1 and β2, and the other between β2 and αD. There are a limited number of contacts between the two domains, which are linked by only a single β-strand (β1).

In the crystal lattice there is an obvious octamer with approximate dimensions 96 x 96 x 110 Å. The octamer has 42 symmetry and is most conveniently described as being formed from a tetramer of dimers (Fig. 1C). The solvent-accessible surface area of an isolated monomer, calculated using the programme AREAIMOL with a probe radius of 1.4 Å (17), is 9400 Å2. On formation of the octamer, 3200 Å2 (34%) of the solvent-accessible surface is buried per monomer. Of the three distinct molecular two-fold axes in the octamer, two are crystallographic and the other is non-crystallographic. Adjacent monomers related by the crystallographic two-fold axes in the octamer form a dimer that buries 2100 Å2 (22%) of the solvent-accessible surface per monomer (Fig. 1D). Adjacent dimers related by the non-crystallographic two-fold axis in the octamer form a dimer-dimer

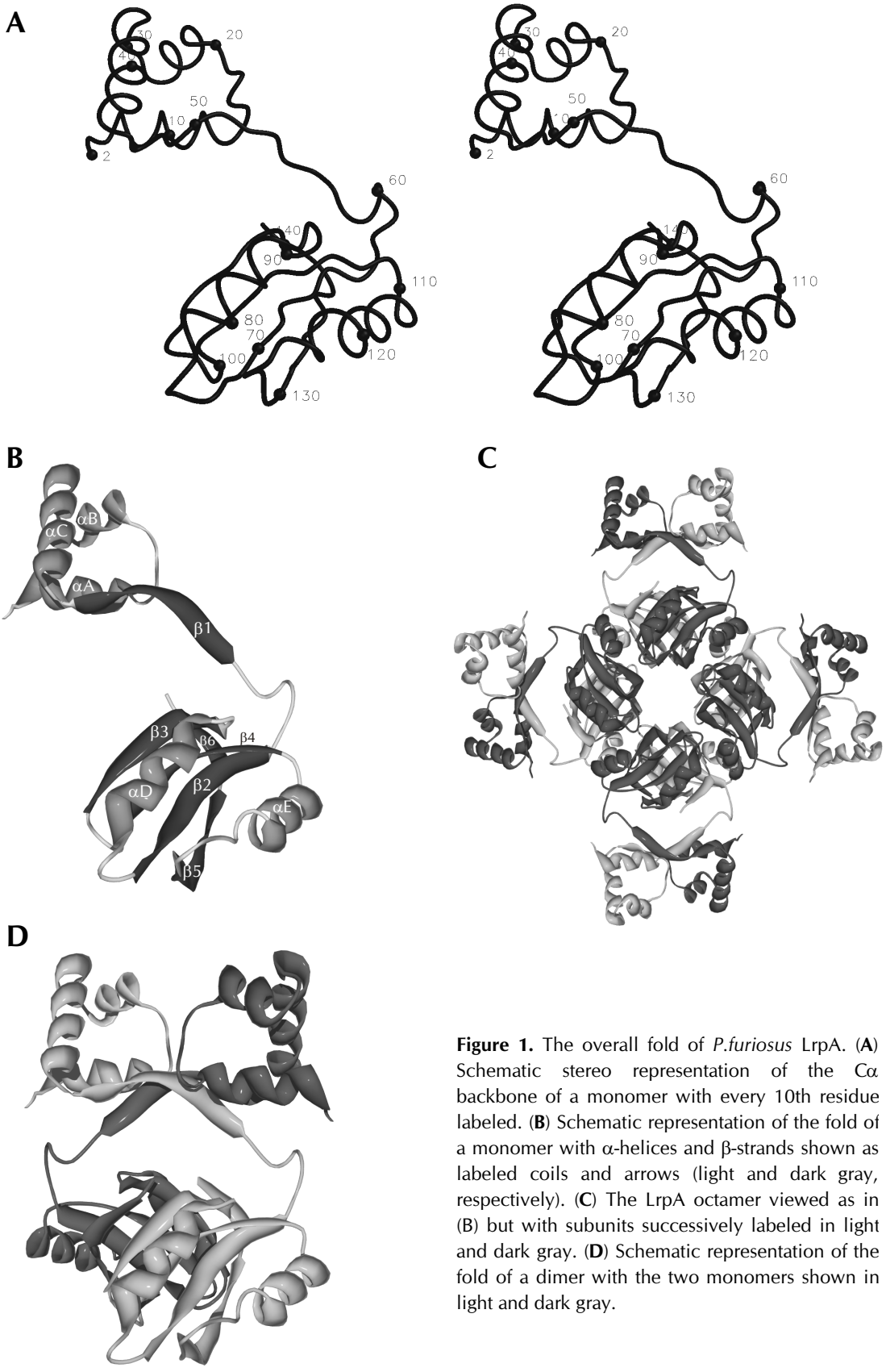



**Figure 1.** The overall fold of *P.furiosus* LrpA. (**A**) Schematic stereo representation of the Cα backbone of a monomer with every 10th residue labeled. (**B**) Schematic representation of the fold of a monomer with α-helices and β-strands shown as labeled coils and arrows (light and dark gray, respectively). (**C**) The LrpA octamer viewed as in (B) but with subunits successively labeled in light and dark gray. (**D**) Schematic representation of the fold of a dimer with the two monomers shown in light and dark gray.

interface that buries a further 1100 Å2 (12%) of the solvent-accessible surface per monomer (Fig. 1C). These interfaces are described below.

## Dimer interface

The interactions that form the dimer interface can be divided into three main regions. First, a hydrophobic core is held together by interactions between residues in strands β2, β3, β4 and β5 from each monomer and, in addition, the β-sheets are extended to form five-stranded antiparallel β-sheets by main chain hydrogen bonding of strand β6 to strand β3 in the other monomer. Secondly, extensive hydrogen bonding interactions can be seen in the anti-parallel β-ribbon formed by the β1 strands from both subunits. The third region of contact is hydrophobic in character and is formed between the N- and C-terminal domains of symmetry-related subunits. Specifically, residues from helix αA and the following turn in the N-terminal domain of one monomer interact with the first $3_{10}$ helical turn and residues from strand β1 of the inter-domain β-ribbon in the C-terminal domain of the second monomer.

## Dimer-dimer interface and octamer formation

In forming the octamer, four equivalent dimer-dimer contacts are formed that are hydrophobic in nature. For any given monomer, contacts between subunits related by a molecular four-fold axis involve helix αE and strand β5 with the $3_{10}$ helical turn between strand β2 and helix αD and the turn between strands β3 and β4. Helix αE makes a further contact with a non-crystallographically related monomer in an adjacent dimer via interactions with the C-terminal residues.

## Structure comparison

The structure of LrpA was compared with those of all the proteins in the Protein Data Bank (PDB) (2) using the program PROTEP (10). Although there was no overall match with the entire structure of LrpA, a number of hits were observed for the C-terminal domain. The three best hits were found with the N-terminal domains of the archaeal DNA polymerase B enzymes from *Thermococcus gorgonarius* (12) and *Desulfurococcus* strain Tok (37), which have been proposed to bind RNA (37), and the ribosomal protein S6 from the small ribosomal subunit of *Thermus thermophilus* (18), which, in conjunction with the S18 protein, binds the ribosomal 16S RNA (29). The structural motif common to these proteins consists of a four-stranded anti-parallel β-sheet with two α-helices packed on one side. This architecture is present in a number of small single-stranded RNA-binding modules

including S6, which, within this structural motif, contain conserved sequences known as RNP1 and RNP2. The structure of the RNA-binding domain (RBD) of the U1A spliceosomal protein complexed with an RNA hairpin is representative of an RBD-RNA complex (26) and shows that the sequence motifs are positioned on the first and third β-strand of the βαββαβ fold, making strong contacts with the bound RNA molecule. Such conserved sequences are not present in LrpA and, given that the surface of the β-sheet that binds the RNA in an RBD is involved in the formation of the dimer interface in LrpA, it is unlikely that LrpA binds RNA molecules via its C-terminal domain. A βαββαβ fold similar to that in the C-terminus of LrpA has also been observed in the C-terminal DNA-binding domain of bovine papillomavirus-1 E2, whose structure has been solved at 1.7 Å bound to its smoothly bent DNA target (11). Like the C-terminal domain of LrpA, in E2 an equivalent region of the subunit surface is also involved in the formation of a dimer, but the subunit-subunit contacts are predominantly hydrophilic. In E2 this C-terminal domain is also involved in binding DNA through interactions between the first helix of the βαββαβ fold of each monomer and the major groove of the DNA double helix. In contrast, the equivalent region in LrpA is involved in dimer-dimer interaction within the octamer, further demonstrating the versatility of this structural motif.

The structure of *P.furiosus* LrpA reveals the presence of an HTH motif between residues 21 and 45 in the N-terminal domain, formed by helices αB and αC. This N-terminal region appears to form a distinct 'headpiece' to the molecule. The HTH motif in LrpA can be superimposed on those from the *E.coli* catabolite activator protein (20) and the tryptophan repressor (30) with a root mean square deviation (RMSD) of 0.51 and 1.87 Å, respectively, for the 20 αC atoms of the motif (residues 21-40 in LrpA). The use of distinct HTH-containing 'headpiece' domains to bind the DNA has been observed in a number of prokaryotic transcriptional regulators including the lac repressor (9).

## Mutational analysis of Lrp/AsnC family proteins

The structure of *P.furiosus* LrpA is the first to be solved for a member of the Lrp/AsnC family. This structural information can be used to make a structure-based sequence alignment (Fig. 2A), allowing comparisons to be made between members of the family, and facilitating the interpretation of biochemical data that exist for *E.coli* Lrp in the light of the current *P.furiosus* LrpA model. This close sequence similarity within the Lrp family combined with mutation studies carried out on *E.coli* Lrp (27) allows us to investigate the location of both the effector and DNA binding sites. *Escherichia coli* Lrp has been randomly mutated and the resulting mutants tested on the basis of their effects on

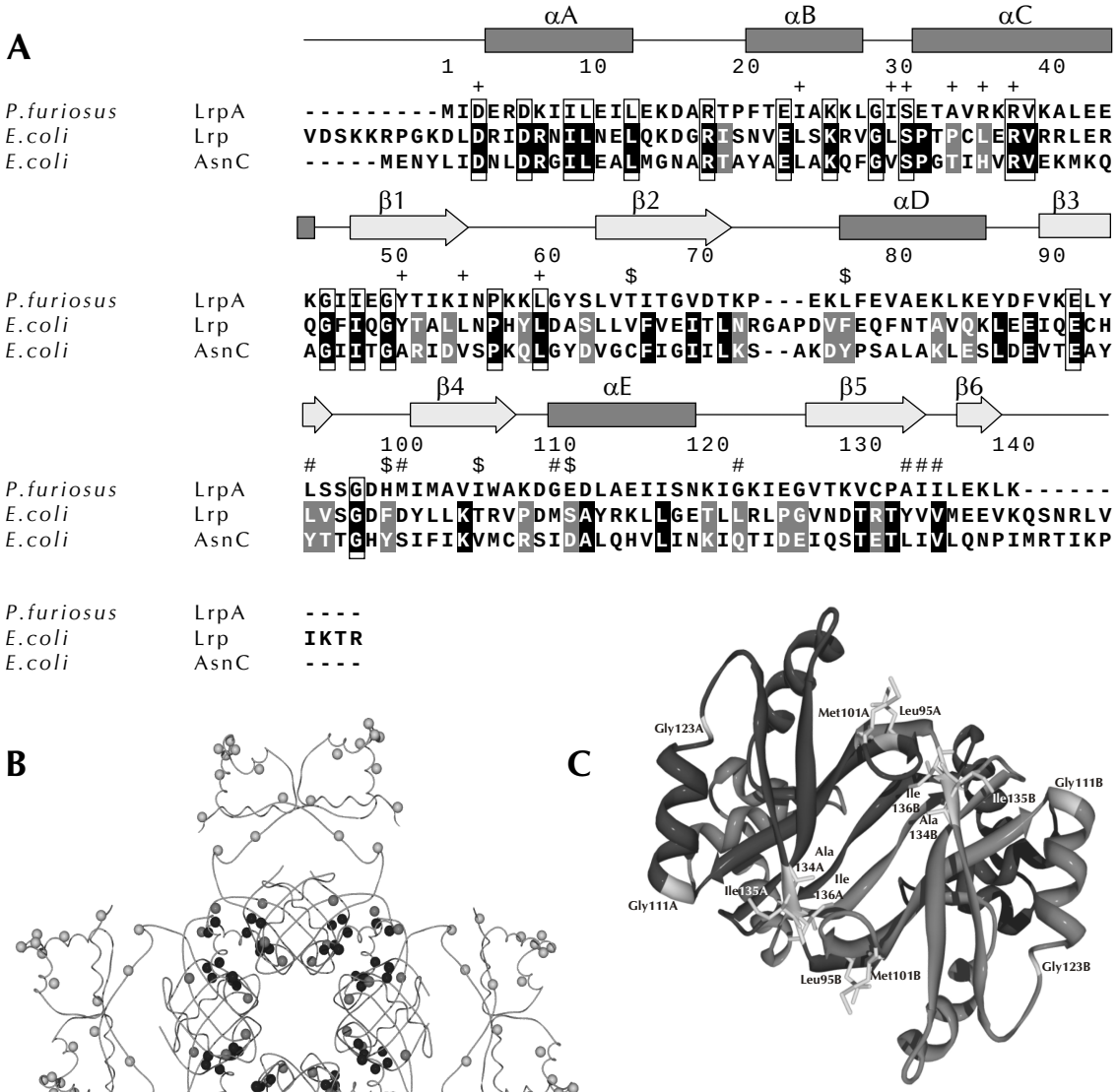



**Figure 2.** Sequence alignment and location of DNA binding, activation and leucine response mutations. (**A**) Structure-based multiple alignment of Lrp/AsnC family sequences. Elements of secondary structure in LrpA are shown as labeled cylinders (α-helices) and arrows (β-strands). Sequences are aligned from *P.furiosus* LrpA, *E.coli* Lrp and *E.coli* AsnC. Residues that are conserved across all three sequences have been boxed. Those residues that are conserved between Lrp and AsnC are shaded black. Those residues that are switched from a hydrophobic side chain in Lrp to a hydrophilic side chain in AsnC are shaded gray. The positions of the *E.coli* Lrp DNA binding mutants, activation mutants and leucine response mutants are indicated by the symbols +, $ and #, respectively. (**B**) The LrpA octamer as shown in Figure 1C but with all subunits displayed as coils. The Cα atoms of DNA binding mutants, activation mutants and leucine response mutants are shown in space-filling style in white, gray, and black, respectively. (**C**) An LrpA dimer viewed along its two-fold axis (i.e. rotated 90° around the *x*-axis with respect to Figure 1D). The monomers are shown in gray and dark gray, and the equivalent residues in LrpA to those identified in *E.coli* Lrp as leucine response mutants are shown in white. The residue sequence numbers are those of LrpA.

expression of *ilvIH*, one of the operons regulated positively by Lrp (28). The *ilvIH* operon encodes an enzyme involved in the biosynthesis of leucine, valine and isoleucine, and expression of this operon is repressed when cells are grown in the presence of leucine. Mutant strains that were resistant to the repressive effects of leucine were termed leucine-response mutants. Those mutants for which binding to *ilvIH* DNA *in vitro* was markedly reduced were termed DNA-binding mutants. A further class of mutants that had low *ilvIH* expression *in vivo* but apparently normal DNA binding *in vitro* were termed activation mutants, owing to their inability to activate transcription. The positions of these mutations have been modelled onto the LrpA structure and provide insights into the patterns of effector molecule and DNA binding. These are described in the following sections.

*Analysis of the pattern of leucine-response mutants and location of the effector-binding site in the Lrp family.* A total of seven *E.coli* Lrp mutants were isolated that were resistant to the repressive effects of leucine (Leu107*, Asp113*, Met123*, Leu135*, Tyr146*, Val147* and Val148*; in the following discussion *E.coli* Lrp and AsnC sequence numbers are denoted by * and ', respectively, and correspond to SwissProt entries P19494 and P03809). When these mutations are modelled onto the structure of the LrpA octamer (the equivalent LrpA residues are Leu95, Met101, Gly111, Gly123, Ala134, Ile135 and Ile136) all seven residues are located at subunit interfaces (Fig. 2B and C). Five out of the seven are found to be clustered in a single region across the dimer interface. The remaining two residues, Met123*/Gly111 and Leu135*/Gly123, are located close to the positions of the other five mutations but in adjacent dimers of the octamer rather than within the same dimer. It is possible, therefore, that effector binding may influence formation of larger multimeric species through additional interaction with residues Met123*/Gly111 and Leu135*/Gly123, although it cannot be ruled out that mutation of these residues may have long-range effects upon leucine binding at the remote site on the dimer interface.

In addition to the above study we attempted to locate possible effector binding sites through comparison of the *E.coli* Lrp and AsnC sequences. We reasoned that if AsnC and Lrp bound their respective effectors at a similar site, such a site might utilize conserved interactions possibly involving charged residues on the protein and the common amino and carboxyl moieties of the effector. In addition, we might anticipate that the specificity for asparagine or leucine displayed by AsnC and Lrp, respectively, might rely on a binding pocket for the side chain of the effector, which would switch in nature from hydrophilic to hydrophobic. Charged residues conserved between Lrp and AsnC (such conservation need not necessarily extend to LrpA since at present there is no evidence that *P.furiosus* LrpA is affected by leucine or any other molecule), and those residues that are hydrophobic in Lrp

but hydrophilic in AsnC, were both plotted onto the LrpA structure. Three sites of conserved charged residues designated A, B and C, which could form possible effector molecule binding sites, were located (A, Asp12*/Asp7', Asp15*/Asp10' and Arg47*/Arg42'; B, Arg27*/Arg22' and Glu32*/Glu27'; C, Glu104*/Glu97' and Lys117*/Lys110'). Close to each of these sites, a residue that switched in nature between hydrophobic and hydrophilic could also be observed (A, Pro43*/Thr38'; B, Ile28*/Thr23'; C, Leu63*/Asp58'). The three sites are all situated within the N-terminal domain or immediately adjacent to it. Thus, if one of the sites represents the binding pocket for an effector molecule then it may well be that effector binding could influence the conformation of the N-terminal domain and, therefore, the relative positions of the DNA-binding helices. However, none of the three sites suggested by the sequence/structure alignment overlaps with that proposed on the basis of the leucine response mutant analysis. The sequence analysis that we have carried out does not preclude the use of main chain atoms or hydrophilic side chains in the binding of the amino and carboxyl groups on the amino acid effector molecule and, thus, these sites may be of little significance. However, it is possible that the previously observed leucine response mutants may have long-range effects and that mutation of residues in the true site is lethal and hence went undetected.

*Analysis of DNA binding and activation mutants.* A representation of the electrostatic surface charge potential of LrpA as computed by GRASP (23) reveals the residues in the recognition helix (αC) of the HTH motif to be predominantly positively charged. In contrast, helix αD in the C-terminal domain, which is analogous to the DNA-binding helix in the bovine papillomavirus-1 E2 DNA-binding domain, is predominantly negatively charged, whereas in bovine papillomavirus-1 E2 it is predominantly positively charged. This implies that *P.furiosus* LrpA is much more likely to bind DNA by interactions between the recognition helices of the HTH motif and the DNA. Analysis of the positions of the residues in the *E.coli* Lrp DNA binding mutants on the structure of LrpA reveals that, of the 10 mutants (Asp12*, Leu33*, Leu39*, Ser40*, Pro43*, Leu45*, Arg47*, Tyr60*, Leu64* and Leu69* in the *E.coli* Lrp sequence, equivalent to Asp3, Ile24, Ile30, Ser31, Ala34, Arg36, Arg38, Tyr51, Ile55 and Leu60 in the LrpA sequence), six are positioned in the HTH motif. Three of these are on the recognition helix αC (Pro43*, Leu45*, Arg47*) and are likely to be directly involved in DNA binding. The remaining four all lie close to the HTH motif and could disturb DNA recognition through long-range conformational effects. Analysis of the positions of the five activation mutants (Val75*, Phe89*, Phe112*, Thr118* and Ser124* in the *E.coli* Lrp sequence, equivalent to Thr66, Leu77, His100,

Ile106 and Glu112 in the LrpA sequence) on the structure of LrpA did not reveal any obvious clustering of residues.

## LrpA-DNA complex model

The two recognition helices (αC) of the HTH motifs in the dimer are separated by 34 Å, which corresponds well to the distance between adjacent turns of the major groove in B-form DNA. In contrast, the helices in the C-terminal domain, which are analogous to the DNA-binding helices in the bovine papillomavirus-1 E2 DNA-binding domain, are separated by 42 Å. Thus, we have modelled straight B-form DNA onto the surface of the LrpA dimer containing the two recognition helices, such that the two-fold axis of the dimer is coincident with a local two-fold in the DNA, and the two recognition helices access adjacent turns of the major groove (Fig. 3). However, inspection of the interactions within this complex suggests that they may not be optimal, particularly when compared with the structures of protein-DNA complexes that utilize the HTH motif. It has been demonstrated that *E.coli* Lrp can induce bending of DNA upon DNA binding (5), and recent studies have suggested that *P.furiosus* LrpA may also be able to bend DNA (3). Thus,

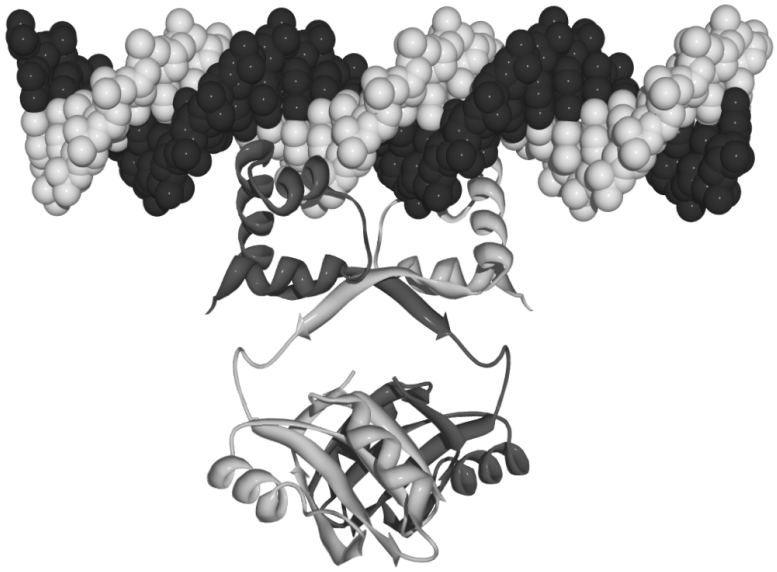


**Figure 3.** Modeling of DNA binding by LrpA. A straight piece of B-form DNA modelled onto an LrpA dimer such that the two-fold axis of the dimer is coincident with a local two-fold in the DNA and the two recognition helices access adjacent turns of the major groove.

it may be that in the true complex, DNA binding would involve interaction with a bent rather than straight DNA. Alternatively, it is also possible that upon binding DNA some small rearrangement of the recognition helices occurs, as there may well be some flexibility

between the N- and C-terminal domains. Consistent with this idea, there are only limited contacts between the two domains of LrpA, and this latter point is given some further support by a superposition of the two molecules in the asymmetric unit, which gives an RMSD of 0.6 Å, whereas the independent superpositions of their N- and C-terminal domains gives RMSD values of 0.5 and 0.4 Å, respectively.

### Thermostability

Differential scanning calorimetry measurements of the LrpA protein from *P.furiosus* have shown that it is extremely thermostable, with a melting temperature, Tm, of 111.5°C (A.B.Brinkman and J.van der Oost, unpublished data). Previous studies have suggested that an increase in the extent of ion pair networks is a frequently observed feature of proteins from hyperthermophiles (33). An analysis of the LrpA structure reveals there to be only 0.057 ion pairs per residue (using a distance limit of 4.0 Å). This is considerably lower than that observed for other hyperthermophilic proteins (33). The analysis is limited somewhat by the relatively low resolution of the LrpA structure and by the absence of a number of charged side chains on the surface of LrpA, which are disordered in the crystal but are presumably ordered upon binding DNA or in interactions with other transcriptional components such as RNA polymerase. Thus, a proper structural understanding of the thermostability of LrpA must await the determination of a comparative structure from a mesophilic member of the Lrp/AsnC family.

## *DISCUSSION*

The work on LrpA presented here describes the first structure of a member of the Lrp/AsnC family of proteins and one of the first structures of a transcriptional regulator from an archaeal source. It reveals a striking octameric assembly formed from a tetramer of dimers. Analysis of the structure and comparisons with sequences of other Lrp/AsnC family members has confirmed the presence of an N-terminal HTH motif and its likely role in DNA binding. In addition, this study has highlighted a potential effector binding site on LrpA via interpretation of mutational analysis of its *E.coli* homologue, and three further potential sites through sequence analysis. The first site appears to straddle the dimer and dimer-dimer interfaces in the octamer and suggests a possible role in effecting the multimeric state of the protein, whereas the locations of the other three possible sites

suggest the potential to influence the conformation of the N-terminal domain and, therefore, the relative positions of the DNA-binding helices.

Gel retardation experiments show that LrpA binds to the *lrpA* promoter region as a single species (3). Whilst chemical cross-linking analysis has suggested the presence of an LrpA tetramer in the protein-DNA complex (3), solution studies (31) and the crystal structure have shown LrpA to exist as an octamer. Further experiments are required to resolve the discrepancy in the molecular sizes determined in these studies. One can speculate that multimerization of the LrpA dimer contributes to stabilizing the DNA-protein complex *in vivo*. At present it is not clear whether LrpA interacts with a single or a multiple operator (3), but by comparison, *E.coli* Lrp has been demonstrated to bind cooperatively to adjacent operators (5). The generally eukaryal-like and hence multi-component nature of the transcriptional machinery observed in the archaea (1) prompts the suggestion that LrpA may also interact with other proteins to form macromolecular complexes during transcriptional regulation. Interaction with DNA-bound TATA binding protein (TBP) (32) has been proposed for the C-terminal domain of papillomavirus-1 E2 protein, which shares the βαββαβ fold topology with the C-terminal domain of LrpA. Future biochemical and crystallographic analyses will address the intriguing possibility that LrpA is involved in the formation of larger multimeric macromolecular complexes *in vivo*.

## EXPERIMENTAL PROCEDURES

Crystals of LrpA were grown by the hanging-drop vapour diffusion method from buffered ammonium sulfate solutions, at both basic and acidic pH, over a range from 4 to 9 as described elsewhere (31). The crystals belong to space group $I4_122$, with cell dimensions a = b = 104.5 Å and c = 245.1 Å, with one dimer in the asymmetric unit and a $V_m$ of 5.2 Å$^3$/Da corresponding to a high solvent content of 70% (19). The crystals had a $d_{min}$ of 2.9 Å at the CLRC Daresbury synchrotron source.

The structure was solved using MIR techniques with two isomorphous derivatives. The first of these was prepared by co-crystallizing the protein in the presence of 0.1 mM ethyl mercuri phosphate (EMP). Data were collected at room temperature on the native and derivative crystals to 4.0 Å on a Mar345 image plate mounted on a Rigaku RU200 X-ray generator. Data were processed using the DENZO suite of programs (25) and subsequently handled using CCP4 software (CCP4, 1994). The Patterson function for this derivative was readily soluble, giving two heavy atom sites, one arising from each of the

two monomers in the asymmetric unit. The two heavy atom sites were refined and a preliminary phase set calculated using MLPHARE (24). A second derivative was produced by soaking a native crystal for 2 h in mother liquor containing 1 mM potassium tetra-cyanoplatinate ($K_2Pt(CN)_4$). Two heavy atom sites, one arising from each of the two monomers in the asymmetric unit, were found by difference Fourier methods, and the derivative was subsequently refined and an improved phase set with an overall figure of merit of 0.53 (acentric 0.51, centric 0.61) was calculated from the two derivatives.

| | **Native** | **Native (cryo)** | **Hg[a]** | **Hg[a] (cryo)** | **Pt[b]** | **Pt[b] (cryo)** |
|---|---|---|---|---|---|---|
| Resolution (Å) | 4.0 | 2.9 | 4.0 | 3.5 | 4.5 | 3.8 |
| No. of observed reflections | 16996 | 30436 | 10680 | 25216 | 14757 | 13699 |
| No. of unique reflections | 5846 | 13804 | 5707 | 7395 | 4200 | 6172 |
| Completeness (%) | 98.3 (99.3) | 94.7 (96.2) | 94.8 (96.2) | 87.0 (92.6) | 99.4 (100) | 91.9 (94.5) |
| $R_{merge}$ (%)[c] | 9.2 (34.2) | 6.5 (34.6) | 8.6 (33.3) | 6.2 (14.9) | 11.1 (28.8) | 8.5 (21.1) |
| Reflection intensities $I/\sigma I > 3$ (%) | 76.8 (47.5) | 75.5 (35.4) | 69.3 (35.9) | 74.2 (30.5) | 65.3 (46.7) | 69.6 (51.1) |
| $R_{iso}$ (%)[d] | | | 15.1 | 18.9 | 13.3 | 29.1 |
| No. of heavy atom sites (monomer) | | | 1 | 1 | 1 | 1 |
| Phasing power (acentric/centric)[e] | | | 2.40/1.89 | 1.26/0.90 | 0.96/0.75 | 1.53/1.09 |
| $R_{cullis}$ (acentric/centric)[f] | | | 0.57/0.48 | 0.79/0.77 | 0.87/0.77 | 0.74/0.70 |

**Table I.** Data proceccing and heavy atom statistics. The data for the highest resolution shells are given in parenthesis. [a]Hg, ethyl mercury phosphate ($C_2H_5HgPO_4$). [b]Pt, potassium tetra-cyanoplatinate ($K_2Pt(Cn)_4$). [c]$R_{merge} = \Sigma|I - <I>|\Sigma I$, where $I$ is the integrated intensity of a given reflection. [d]$R_{iso} = \Sigma|F_{PH} - FP|\Sigma F_P$, where $F_{PH}$ and $F_P$ are the derivative and native structure factor amplitudes. [e]Phasing power = (r.m.s. heavy atom structure factor)/(r.m.s. lack of closure). [f]$R_{cullis}$ = (r.m.s. lack of closure)/(r.m.s. isomorphous difference)

In order to enhance radiation stability and hence improve the resolution of the structural analysis, the crystals were cryoprotected in mineral oil by removing them from the hanging drops with a cryo loop, removing excess precipitant using absorbent dental points, dragging the loop through an oil reservoir and placing it in the cryo stream. This enabled a 2.9 Å resolution native data set and 3.8 Å resolution $K_2Pt(CN)_4$ derivative data set to be collected. Both data sets were collected on a Mar image plate detector on station 7.2 at the CLRC Daresbury Laboratory. Finally, a 3.5 Å resolution EMP derivative data set was collected on a Mar345 image plate mounted on a Rigaku RU200 X-ray generator. The cryo-cooling produced a 3% shrinkage of the cell in the *a* and *b* dimensions (101.3 Å) but little change in *c* (245.4 Å) (Table I).

An electron density map was calculated to 3.5 Å resolution and improved by solvent flattening and two-fold non-crystallographic symmetry averaging using the program DM (6). The resolution of the map was improved by phase extension to 2.9 Å and showed

clearly identifiable regions of regular secondary structure. The map was skeletonized using the program MAPMAN (14) and a polyalanine model for a single subunit was constructed with the program O (13). This model was rotated approximately into the electron density for the second subunit in the asymmetric unit using the program PDBSET, and its position refined using rigid body refinement in O. Subsequently, the sequence was fitted into the model where it could be unambiguously assigned, and when 80% of the total number of side chains had been determined with confidence the structure was submitted to maximum likelihood refinement using the program REFMAC (21). Iterative cycles of phase combination of the partial structure phases and those from the heavy atom derivatives, model building and refinement were used to construct a complete model representing 280 out of 282 expected residues. NCS restraints were applied between the subunits and an overall average *B*-factor (estimated from a Wilson plot of the data) of 62 $Å^2$ was used. The electron density for a total of 18 side chains in the A and B subunits of the asymmetric unit was not observed in the final electron density map and was all truncated at the β-carbon. The final model has an *R*-factor of 31.3% ($R_{free}$, 38.2%; (4)) for all data in the resolution range 20-2.9 Å. The model has good stereochemistry, with values for the RMSD from standard values of the bond lengths and angles of 0.012 Å and 3.0°, respectively. Model geometry was analyzed using the program PROCHECK (16). A Ramachandran plot of the model shows all non-glycine residues inside the normally allowed regions (87.6 and 12.4% in the most favored and additional allowed regions, respectively) and examination of χ1-χ2 plots for all residue types showed no side chains in unfavourable conformations.

The coordinates and structure factor amplitudes have been submitted to the RCSB PDB; code 1I1G.

## ACKNOWLEDGEMENTS

We thank BBSRC, Wellcome Trust and The New Energy and Industrial Development Organization for their support. J.B.R. is the Royal Society Olga Kennard Fellow. The Krebs Institute is a BBSRC designated Biomolecular Sciences Centre, and its structural studies group is a member of the BBSRC North of England Structural Biology Centre. Part of this research was supported by grant 700-35-101 of the Council for Chemical Sciences (CW) of The Netherlands Organisation for Scientific Research (NWO).

## REFERENCES


1. Bell, S. D., S.P. Jackson. 1998. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends in Microbiology. 6:222-228.
2. Bernstein, F.C., T.F. Koetzle, G.J. Williams, E.E. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures J Mol Biol. 112:535-42.
3. Brinkman, A.B., I. Dahlke, J.E. Tuininga, T. Lammers,. V. Dumay, E. de Heus, J.H.G. Lebbink, M. Thomm, W.M. de Vos, J. van Der Oost. 2000. An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated J Biol Chem. 275:38160-38169.
4. Brunger, A.T. 1992. Free R-value - a novel statistical quantity for assessing the accuracy of crystal structures Nature. 355:472-475.
5. Calvo, J.M., and R.G. Matthews 1994. The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiol Rev. 58:466-490.
6. Cowtan, K.D. 1994. An automated procedure for phase improvement by density. Joint CCP4 and ESF-EACBM newsletter on protein crystallography. 31: 34-38
7. Cui, Y., Q. Wang, G.D. Stormo, J.M. Calvo. 1995. A consensus sequence for binding of Lrp to DNA J Bacteriol. 177:4872-4880.
8. de Wind, N., M. de Jong, M. Meijer, and A.R. Stuitje 1985. Site-directed mutagenesis of the *Escherichia coli* chromosome near *oriC*: identification and characterization of *asnC*, a regulatory element in *E.coli* asparagine metabolism. Nucleic Acids Res. 13:8797-811.
9. Friedman, A.M., T.O. Fischmann, and T.A. Steitz 1995. Crystal structure of Lac repressor core tetramer and its implications for DNA looping. Science. 268:1721-7.
10. Grindley, H.M., P.J. Artymiuk, D.W. Rice, and P. Willett 1993. Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. J Mol Biol. 229:707-21.
11. Hegde, R.S., S.R. Grossman, L.A. Laimins, and P.B. Sigler 1992. Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. Nature. 359:505-12.
12. Hopfner, K.P., A. Eichinger, R.A. Engh, F. Laue, W. Ankenbauer, R. Huber, and B. Angerer 1999. Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. Proc Natl Acad Sci U S A. 96:3600-5.
13. Jones, T.A., J.Y. Zou, S.W. Cowan, and M. Kjeldgaard 1991. Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A. 47:110-9.
14. Kleywegt, G.J., and T.A. Jones 1994. Halloween...masks and bones, p. 59-66. In S. Bailey, R. Hubbard, and D. Waller (eds), From first map to final model. SERC Daresbury Laboratory, Warrington, UK.
15. Kölling, R., and H. Lother 1985. AsnC: an autogenously regulated activator of asparagine synthetase A transcription in *Escherichia coli*. J Bacteriol. 164:310-5.
16. Laskowski, R.A., M.W. MacArthur, D.S. Moss, and J.M. Thornton 1993. PROCHECK: a program to check the stereochemical quality of protein structures J Appl Crystallogr. 26:283-291.
17. Lee, B., and F.M. Richards 1971. The interpretation of protein structures: estimation of static accessibility. J Mol Biol. 55:379-400.

18. Lindahl, M., L.A. Svensson, A. Liljas, S.E. Sedelnikova, I.A. Eliseikina, N.P. Fomenkova, N. Nevskaya, S.V. Nikonov, M.B. Garber, T.A. Muranova. 1994. Crystal st ructure of the ribosomal protein S6 from *Thermus thermophilus*. EMBO J. 13:1249-54.
19. Matthews, B.W. 1977. The X-ray structure of proteins, p. 468-477. In H. Neurath, and R. L. Hill (eds), The Proteins, vol. 3. Academic Press, New York.
20. McKay, D.B., and T.A. Steitz 1981. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. Nature. 290:744-9.
21. Murshudov, G.N., A.A. Vagin, and E.J. Dodson 1997. Refinement of macromolecular structures by the maximum-likelyhood method. Acta Crystallogr D Biol Crystallogr. 53:240-255.
22. Newman, E.B., Lin, R. 1995. Leucine-responsive regulatory protein: a global regulator of gene expression in *E. coli*. Ann Rev Microbiol. 49:747-775.
23. Nicholls, A., K.A. Sharp, and B. Honig 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins. 11:281-96.
24. Otwinowski, Z. 1991. Maximum likelyhood refinement of heavy atom parameters, p. 80-86. In W. Wolf, P. R. Evans, and A. G. W. Leslie (eds), Proceedings of the CCP4 study weekend: Isomorphous replacement and anomalous scattering. SERC Daresbury Laboratory, Warrington, UK.
25. Otwinowski, Z., and W. Minor 1997. Processing of X-ray diffraction data collected in oscil lation mode Meth Enzymol. 276:307-326.
26. Oubridge, C., N. Ito, P.R. Evans, C.H. Teo, and K. Nagai 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature. 372:432-8.
27. Platko, J.V., J.M. Calvo. 1993. Mutations affecting the ability of *Escherichia coli* Lrp to bind DNA, activate transcription, or respond to leucine. J Bacteriol. 175:1110-1117.
28. Platko, J.V., D.A. Willins, and J.M. Calvo 1990. The *ilvIH* operon of *Escherichia coli* is positively regulated. J Bacteriol. 172:4563-70.
29. Powers, T., and H.F. Noller 1995. Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. RNA 1:194-209.
30. Schevitz, R.W., Z. Otwinowski, A. Joachimiak, C.L. Lawson, and P.B. Sigl er 1985. The three-dimensional structure of Trp repressor. Nature. 317:782-6.
31. Sedelnikova, S.E., S.H. Smits, P.M. Leonard, A.B. Brinkman, J. van der Oost, J.B. Rafferty, and D.W. Rice 2001. Crystallization and quaternary structure analysis of an Lrp-like regulatory protein from the hyperthermophile *Pyrococcus furiosus* Acta Crystallogr D Biol Crystallogr. 57:886-8.
32. Steger, G., J. Ham, O. Lefebvre, and M. Yaniv 1995. The bovine papillomavirus 1 E2 protein contains two activation domains: one that inte racts with TBP and another that functions after TBP binding. EMBO J. 14:329-40.
33. Szilagyi, A., and P. Zavodszky 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehens ive survey. Structure Fold Des. 8:493-504.
34. Wang, Q., J.M. Calvo. 1993. Lrp, a global regulatory protein of *Escherichia coli*, binds co-operatively to multiple sites and activates transcription of *ilvIH*. J Mol Biol. 229:306-318.
35. Wang, Q., J. Wu, D. Friedberg, J. Plakto, J.M. Calvo. 1994. Regulation of the *Escherichia coli lrp* gene. J Bacteriol. 176:1831-1839.

36. Willins, D.A., C.W. Ryan, J.V. Platko, J.M. Calvo. 1991. Characterization of Lrp, an *Escherichia coli* regulatory protein that mediates a glo bal response to leucine. J Biol Chem. 266:10768 -10774.
37. Zhao, Y., D. Jeruzalmi, I. Moarefi, L. Leighton, R. Lasken, and J. Kuriyan 1999. Crystal structure of an archaebacterial DNA polymerase. Structure Fold Des. 7:1189 -99.

# Chapter 6

## The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability

Arie B. Brinkman
Stephen D. Bell
Robert Jan Lebbink
Willem M. de Vos
John van der Oost

**Although the archaeal transcription apparatus resembles the eukaryal RNA polymerase II system, many bacterial-like regulators can be found in archaea. Particularly, all archaeal genomes sequenced to date contain genes encoding homologues of Lrp (leucine-responsive regulatory protein). Whereas Lrp-like proteins in bacteria are involved in regulation of amino acid metabolism, their physiological role in archaea is unknown. Although several archaeal Lrp-like proteins have been characterized recently, no target genes apart from their own coding genes have been discovered yet, and no ligands for these regulators have been identified so far. In this study we show that the Lrp-like protein LysM from *Sulfolobus solfataricus* is involved in the regulation of lysine, and possibly also arginine biosynthesis, encoded by the *lys* gene cluster. Exogenous lysine is the regulatory signal for *lys* gene expression and specifically serves as a ligand for LysM by altering its DNA-binding affinity. LysM binds directly upstream of the TFB-responsive element of the intrinsically weak *lysW* promoter, and DNA binding is favored in the absence of lysine, when *lysWXJK* transcription is maximal. The combined *in vivo* and *in vitro* data are most compatible with a model in which the bacterial-like LysM activates the eukaryal-like transcriptional machinery. As with transcriptional activation by *E.coli* Lrp, activation by LysM is apparently dependent on a co-activator, the identity of which remains to be identified.**

*Journal of Biological Chemistry, 2002, in press*

## INTRODUCTION

Since the discovery of archaea as a distinct domain of life, many studies have focussed on archaeal transcription. It has become clear that although archaea resemble bacteria with respect to their cellular and genetic organization, their transcriptional apparatus is fundamentally different from that of bacteria. Their RNA polymerase (RNAP) is much more related to the eukaryal RNAPII system regarding subunit complexity and sequence homology (28). Thus, archaeal RNAP consists of at least 10 subunits in contrast to the five-subunit bacterial RNAP core enzyme. As in eukarya, archaeal transcription initiation is preceded by the binding of the TATA-binding protein (TBP) to a TATA-like sequence called the TATA-box, and subsequent binding of transcription factor B (TFB). Archaeal TBP and TFB are highly homologous to the eukaryal TBP and TFIIB, respectively. However, archaeal TBP is not complexed with TBP-associated factors (TAFs) as in eukarya (45), and there is no evidence that archaeal genomes encode TAF-homologues. The archaeal TATA-box is eight bp in length and is located approximately 25 bp upstream of the start of transcription. Directly upstream of the TATA-box a purine-rich sequence is present, called the TFB responsive element (BRE). The BRE was shown to be an important determinant in directionality of transcription and promoter strength through interaction with a C-terminal helix-turn-helix domain of TFB (6, 32). The TF(II)B-BRE interaction is a conserved feature between archaea and eukarya. Once TBP and TFB are bound to the promoter, RNAP is recruited involving an interaction between the RpoK subunit of RNAP and the N-terminal Zn-ribbon domain of TFB (35).

Although no archaeal homologues of eukaryal TFIIA, TFIIF and TFIIH have been identified, a protein homologous to the N-terminal region of the alpha subunit of eukaryal TFIIE is present in archaea. This archaeal TFE stimulates transcription from promoters with sub-optimal TATA-box sequences, or in cases where TBP is limiting (2, 19). Whereas eukaryal TFIIE is strictly necessary for transcription, archaeal TFE appears to be dispensable for basal transcription *in vitro*, although it may play a stimulatory role in transcription initiation at specific promoters.

Although the basal components of the archaeal and eukaryal transcription machineries are very similar, regulatory proteins do not appear to be conserved between the two domains. Instead, archaeal genomes contain many regulators previously identified only in bacteria, so-called bacterial-archaeal (BA) regulators (3). In particular, homologues of the Lrp/AsnC family of regulators appear to be widely distributed among both bacteria and archaea. Several bacterial as well as archaeal genomes contain up to ten Lrp-like paralogues.

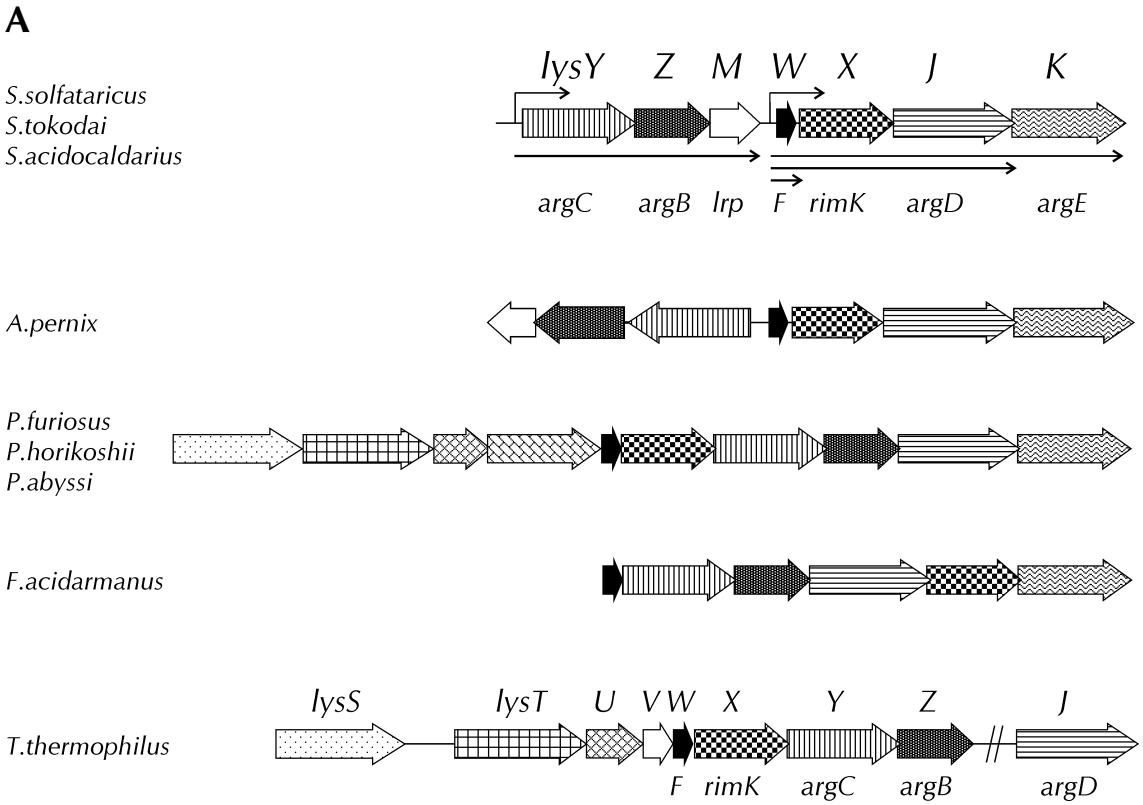
*E.coli* leucine-responsive regulatory protein (Lrp) is the paradigm that has been studied extensively (11, 41). It is a global regulator controlling the expression of up to 75 genes (16, 31). *E.coli* Lrp either represses or activates transcription, the effect of which is sometimes modulated by leucine. The target genes of *E.coli* Lrp encode enzymes that are directly or indirectly related to amino acid metabolism. This also appears to be the case for several specific (non-global) bacterial Lrp-like regulators from different bacteria. In archaea, the exact role of the numerous Lrp-like proteins has not been established. Several archaeal Lrp-like proteins have been characterized recently (4, 9, 15, 40). For two of these proteins, Lrs14 from *S.solfataricus* and LrpA from *P.furiosus*, an *in vitro* regulatory function could be assigned: both showed negative autoregulation independent of any amino acid ligand (4, 9). Moreover, the three-dimensional structure of *P.furiosus* LrpA was determined, providing the structural basis for understanding LrpA-DNA as well as LrpA-ligand interactions (29). However, neither the identity of this ligand nor the role of archaeal Lrp-like proteins in the expression of other genes has been determined.

To provide a suitable model system for analyzing the function of Lrp-like proteins in archaea, we have screened the genome of the hyperthermophilic archaeon *S.solfataricus* for the presence of Lrp-like proteins which function, target, and ligand may be readily predicted. We have identified and characterized the Lrp-like protein LysM, the gene of which is clustered with genes encoding lysine biosynthetic enzymes. We show here that expression of the *lysWXJK* genes is regulated by the presence of lysine in the medium. *In vitro* LysM binds to the *lysW* promoter and binding is favored in the absence of lysine, when *lys* gene expression is maximal. A model is proposed for lysine-modulated activation of transcription through LysM, for the first time indicating that a bacterial-like regulator may activate the eukaryal-like archaeal transcriptional machinery. It appears that Lrp-like proteins are functionally equivalent in the bacterial and archaeal domains, despite the fundamental differences in transcriptional machineries.

## *RESULTS*

### Identification of the lysine biosynthesis gene cluster

In the genome of *Sulfolobus solfataricus* P2 (50) a gene encoding an Lrp-like protein (*lysM*, Sso0157, see Fig. 1A) is present. The *S.solfataricus* LysM protein is 29% and 33% identical to the archaeal Lrp-like proteins Ptr2 from *Methanococcus jannaschii* and LrpA from *Pyrococcus furiosus*, respectively, and 15% and 27% identical to the bacterial Lrp-like proteins



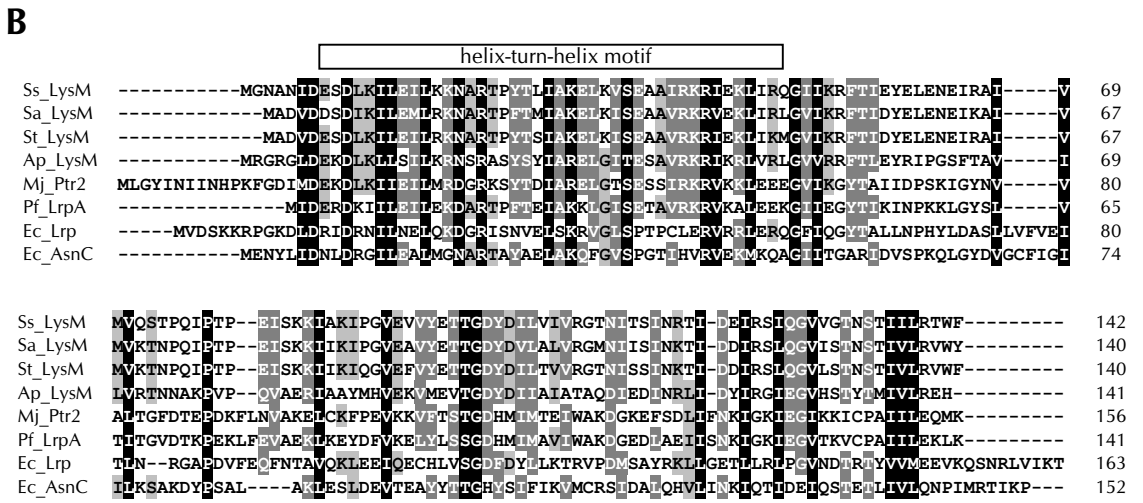
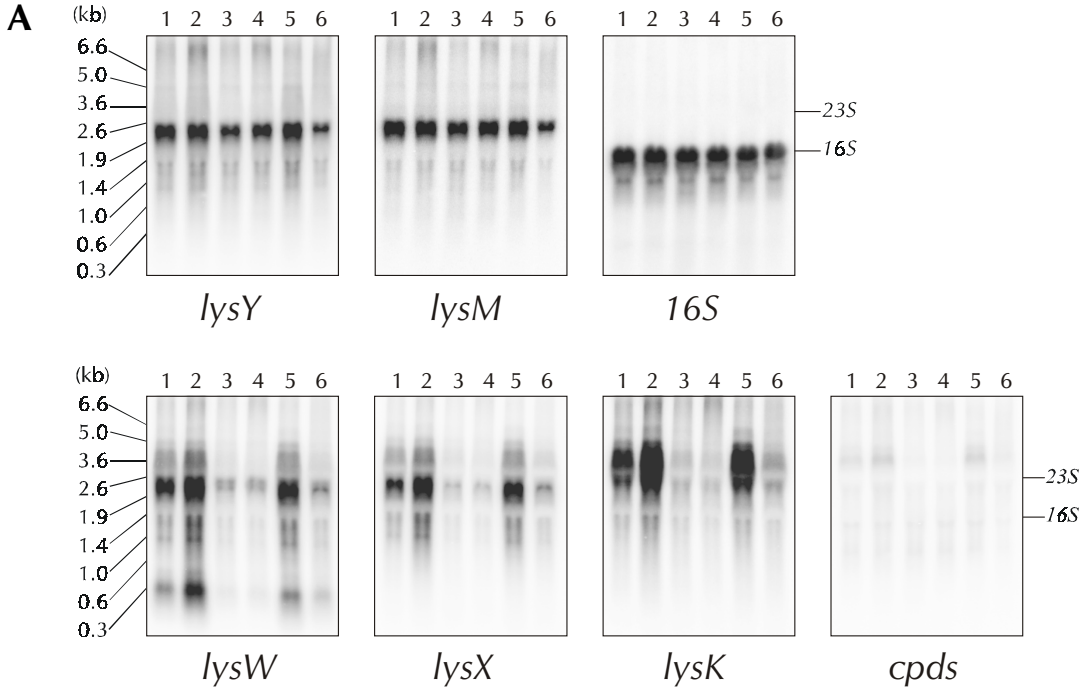



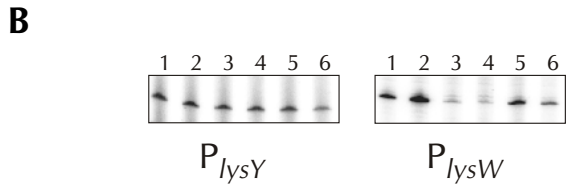

**Figure 1.** (**A**), The *lys* gene cluster of *Sulfolobus solfataricus*, compared to the those of the archaea *Sulfolobus tokodai* (22), *Sulfolobus acidocaldarius* (Q. She, personal communication), *Pyrococcus furiosus* (http://www.genome.utah.edu), *Pyrococcus horikoshii* (24), *Pyrococcus abyssi* (http://www.genoscope.cns.fr), *Aeropyrum pernix* (23), *Ferroplasma acidarmanus* (http://www.jgi.doe.gov/JGI_microbial/html), and the thermophilic bacterium *T.thermophilus* (42). Patterns indicate homology between the encoded proteins. Promoters and length of produced transcripts in *S.solfataricus* are indicated by arrows. (**B**), Multiple alignment of archaeal LysM proteins with other archaeal and bacterial Lrp-like proteins. The helix-turn-helix motif of *Pyrococcus furiosus* LrpA (29) is indicated. Ss_LysM, Sso0157; Sa_LysM, not annotated; St_LysM, Sto0193; Ap_LysM, not annotated; Mj_Ptr2, Q58133; Pf_LrpA, P42180; Ec_Lrp, P19494; Ec_AsnC, P03809.

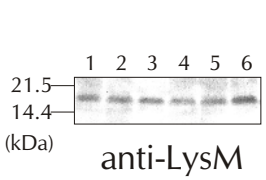
Lrp and AsnC from *Escherichia coli*, respectively (Fig. 1B). In *S.solfataricus* the gene encoding LysM is part of a gene cluster. Four genes of this cluster are homologous to classical arginine biosynthesis genes (*argBCDE*), one is homologous to *Escherichia coli rimK*, encoding a ribosomal protein modification enzyme, and one of the genes is homologous to *Thermus thermophilus orfF*, encoding a small hypothetical protein. Similar gene clusters are also present in at least seven archaeal genomes and one bacterial genome (Fig. 1A). While *argD* of *T.thermophilus* is not clustered with *argBCE*, it is present elsewhere on the genome (39) (see Fig. 1A). The role of the *T.thermophilus* cluster has been studied using gene-disruption of the *argB*, *argC*, *argD*, *orfF* or *rimK* genes, which resulted in lysine auxotrophy (39, 42). Because *T.thermophilus* does not synthesize lysine via the diaminopimelic acid (DAA) pathway, believed to be common to all bacteria, but via α-aminoadipic acid (AAA) as an intermediate (26, 27), it was proposed that the *T.thermophilus orfF-rimK-argCBD* genes are involved in lysine biosynthesis through a modified AAA pathway, in which the conversion of AAA to lysine is similar to the conversion of glutamate to ornithine in the arginine biosynthesis pathway. The cluster was therefore renamed as the *lys* operon (42). Hence, we will refer to the respective *S.solfataricus* gene cluster as the *lys* gene cluster, and we have renamed the genes of the described *S.solfataricus* gene cluster accordingly (see Fig. 1A). The gene encoding the Lrp-like protein LysM is only present within the *lys* clusters of the three *Sulfolobus* species and *A.pernix*. *S.solfataricus* LysM has 74%, 79%, and 44% identity with the *S.acidocaldarius*, *S.tokodai*, and *A.pernix* orthologues, respectively. BLAST and genomic context analysis revealed that no additional genomes present in the current database contain close homologues (≥44% identity) of LysM, suggesting that LysM is restricted to crenarchaea. Since the *lysM* gene is clustered with putative lysine biosynthesis genes of *S.solfataricus*, we hypothesized that LysM could be involved in the regulation of these genes. In analogy, most bacterial Lrp-like proteins are involved in regulation of amino acid metabolism, and their regulatory effect is modulated by one or more amino acids (10).

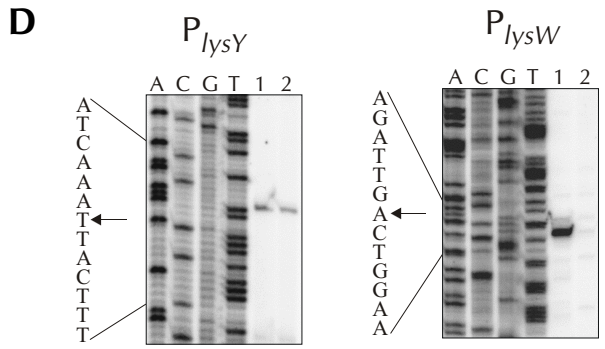
## Expression of the *lys* genes *in vivo*

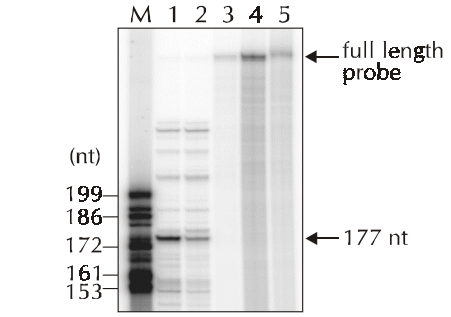
To study *in vivo* expression of the *lys* gene cluster, Northern blotting experiments were performed using total RNA isolated from *S.solfataricus* cells grown to mid-logarithmic phase in defined medium either lacking or containing combinations of amino acids. Probes specific for *lysY* and *lysM* hybridized to identically sized mRNA species of about 2.3 kb expected for a polycistronic mRNA containing *lysYZM*, the level of which appeared to be almost identical under all tested growth conditions (Fig. 2A). Western blotting experiments with a polyclonal antiserum raised against *E.coli*-produced LysM confirmed these results,


A
(kb)
1 2 3 4 5 6
6.6
5.0
3.6
2.6
1.9
1.4
1.0
0.6
0.3
23S
16S
lysY
lysM
16S
lysW
lysX
lysK
cpds


B
1 2 3 4 5 6
PlysY
PlysW


C
1 2 3 4 5 6
21.5
14.4
(kDa)
anti-LysM


D
PlysY
PlysW
A C G T 1 2
ATCAAATTACTTT
AGATTGACTGGAA


E
M 1 2 3 4 5
full length probe
(nt)
199
186
172
161
153
177 nt

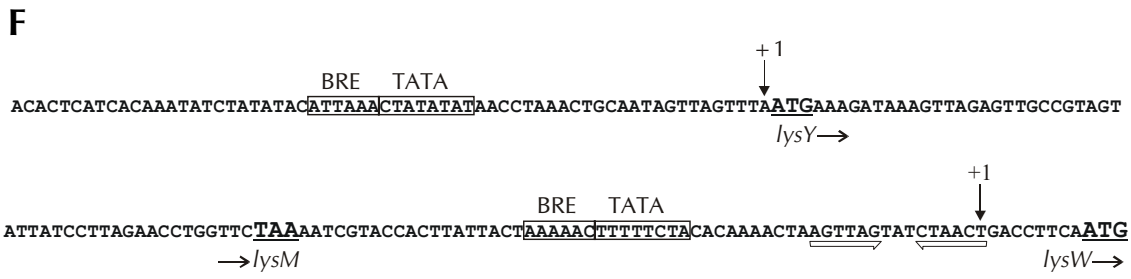

F
+1
BRE TATA
ACACTCATCACAAATATCTATATACATTAAACTATATATAACCTAAACTGCAATAGTTAGTTTAATGAAAGATAAAGTTAGAGTTGCCGTAGT
lysY
+1
BRE TATA
ATTATCCTTAGAACCTGGTTCTAAAATCGTACCACTTATTACTAAAAACTTTTTCTACACAAAACTAAGTTAGTATCTAACTGACCTTCAATG
lysM
lysW

**Figure 2.** (**A**), Northern blotting and (**B**), quantitative primer extension analysis of *lys* transcripts. *S.solfataricus* was grown on defined medium and combinations of amino acids were added. 1, no amino acids; 2, arginine only; 3, lysine only; 4, arginine and lysine; 5, all 20 amino acids except arginine and lysine; 6, all 20 amino acids. Sizes of RNA marker fragments are indicated, as well as the position of 23S and 16S rRNA. (**C**), Western blot analysis of LysM expression, using *S.solfataricus* cell extracts and an antiserum raised against recombinant LysM. 50 µg of *S.solfataricus* extracts was loaded in each lane. Numbering is according to panel A. (**D**), primer extension mapping of *lys* promoters. 1, defined medium (no amino acids added); 2, rich medium (casamino acids and yeast extract added). (**E**), RNase protection analysis of *S.solfataricus* total RNA, to confirm the position of the *lysWXJK* 5′ terminus as detected by primer extension analysis. The expected size of the resulting radiolabeled antisense RNA fragment after RNase digestion is 177-nt (arrow). M, RNA marker fragments (sizes are indicated at the left, additional bands in the marker lane are contaminations); 1, *S.solfataricus* arginine RNA; 2, *S.solfataricus* arginine/lysine RNA; 3, yeast RNA; 4, yeast RNA (no RNase added); 5, untreated full length probe (arrow). (**F**), results of primer extension and RNase protection promoter mapping. Transcriptional start sites are indicated with vertical arrows (+1). Startcodons and stopcodons are indicated by bold and underlined characters. Open arrows indicate inverted repeat elements. The transcription factor B responsive element (BRE), and TATA-boxes are indicated by boxes.

since the LysM concentration was constant under the growth conditions tested in Northern blotting (Fig. 2C). Using probes against *lysW*, *lysX*, and *lysK*, we detected two mRNA species of 2.3 and 3.2 kb, while an additional 0.2 kb transcript was detected only with a probe against *lysW* (Fig. 2A). This suggested that there are three different transcripts, one containing only *lysW*, one containing *lysWXJ*, and one containing *lysWXJK*. Theoretically, probes against *lysW* and *lysX* should also hybridize to the 3.2-kb *lysWXJK* mRNA, however, although this mRNA species was visible on the membrane, it was less abundant compared to the shorter 2.3-kb *lysWXJ* mRNA. Alternatively, the larger 3.2-kb transcript could be the result of initiation at some place within *lysWXJK*, and transcription read-through downstream of *lysK*. To rule out this possibility, we used a probe against the *cpds* gene immediately downstream of *lysK*, which encodes a putative *cis*-polyprenyl diphosphate synthase. Only background levels were obtained using this probe, indicating that the *cpds* gene is not co-transcribed with the *lys* gene cluster, and that *lysK* is the 3' terminal gene of *lys* operon. It is most likely that only a small fraction of mRNAs read through past *lysJ*, which explains the differential band intensity for 3.2 and 2.3 kb mRNAs when probed with *lysW* or *lysX*.

The *lysWXJK* genes are all strongly regulated specifically by the presence of lysine in the medium. Arginine alone or a mixture of 18 amino acids (arginine and lysine omitted) did not affect the abundance of the *lysWXJK* transcript. However, when lysine alone, both lysine and arginine, or all 20 amino acids were present the amount of transcript decreased drastically. Thus, transcription of *lysWXJK* is induced specifically in the absence of lysine. Quantification of the results obtained from Northern blotting and primer extension analysis revealed that transcript levels under induced conditions are over eight-fold higher

compared to that of non-induced conditions. It is unclear why expression levels of this mRNA are below maximum in the absence of any of the amino acids. Taken together, we conclude that the *lys* genecluster is transcribed in two separate polycistronic mRNAs. The transcription of *lysYZM* is almost constitutive, whereas the level of *lysWXJK* mRNA falls drastically whenever lysine is present in the medium.

To determine the 5' end of the two transcripts we performed primer extension analysis. We found that the transcriptional start of *lysY* is located at an adenine preceding the predicted ATG startcodon of *lysY* (Fig. 2D and F). Putative BRE and TATA promoter sequences are present 25 bp upstream from the *lysY* transcriptional start. The transcriptional start of *lysW* is located nine bp upstream from the predicted *lysW* ATG startcodon. Non-canonical BRE and TATA sequences can be recognized 25 bp upstream from the *lysW* transcriptional start. No obvious Shine-Dalgarno sequences are present upstream the *lysY* and *lysW* translational starts. This is not uncommon in *S.solfataricus*; a complete genome analysis showed that that genes at the 5' extremity of putative polycistronic mRNAs often lack ribosome binding sites upstream their translational starts, whereas downstream genes within operons do contain Shine-Dalgarno sequences (52). A detailed sequence analysis of the *lysW* promoter revealed the presence of a perfect 15-bp inverted repeat directly upstream the identified *lysW* transcriptional start that could form a hairpin structure in the mRNA, causing stalling and eventually termination of reverse transcriptase (Fig. 2F). Since this may have obstructed the primer extension analysis we performed RNase protection to confirm the *lysW* transcriptional start detected by primer extension. Assuming that the detected 5' terminus is the true transcriptional start, a 177-nt labeled antisense RNA fragment is expected in RNase protection. As shown in Figure 2E, an RNA fragment with this size is present, confirming that the 5' terminus detected by primer extension is the true transcriptional start. As expected, the intensity of this band decreased when we used RNA isolated from a culture containing lysine.

To verify the variation in transcription levels observed by Northern blotting, we performed quantitative primer extension with the same RNA samples as used for Northern blotting (Fig. 2B). We found the same modulation in transcript levels as shown by Northern blotting experiments. We therefore conclude that transcription of *lysYZM* is driven by the *lysY* promoter, while transcription and regulation of *lysWXJK* occurs from the *lysW* promoter.

## LysM binds to the *lysW* promoter

To study the role of LysM in the regulation of the *lys* gene cluster, we overproduced LysM in *E.coli* in order to facilitate its purification. While significant LysM overproduction was reached, its purification was severely hampered by the tendency of LysM to precipitate irreversibly from the cell extract at a pH lower than 8.0 or in the presence of several salts like $MgCl_2$, NaCl, KCl, or $(NH_4)_2SO_4$. However, using both anion exchange chromatography and a heat-incubation we were able to purify the recombinant LysM to homogeneity, as judged from SDS-PAGE analysis (Fig. 3A). The purified LysM protein was used in electrophoretic mobility shift assays (EMSAs), to determine whether it binds to the mapped promoters.

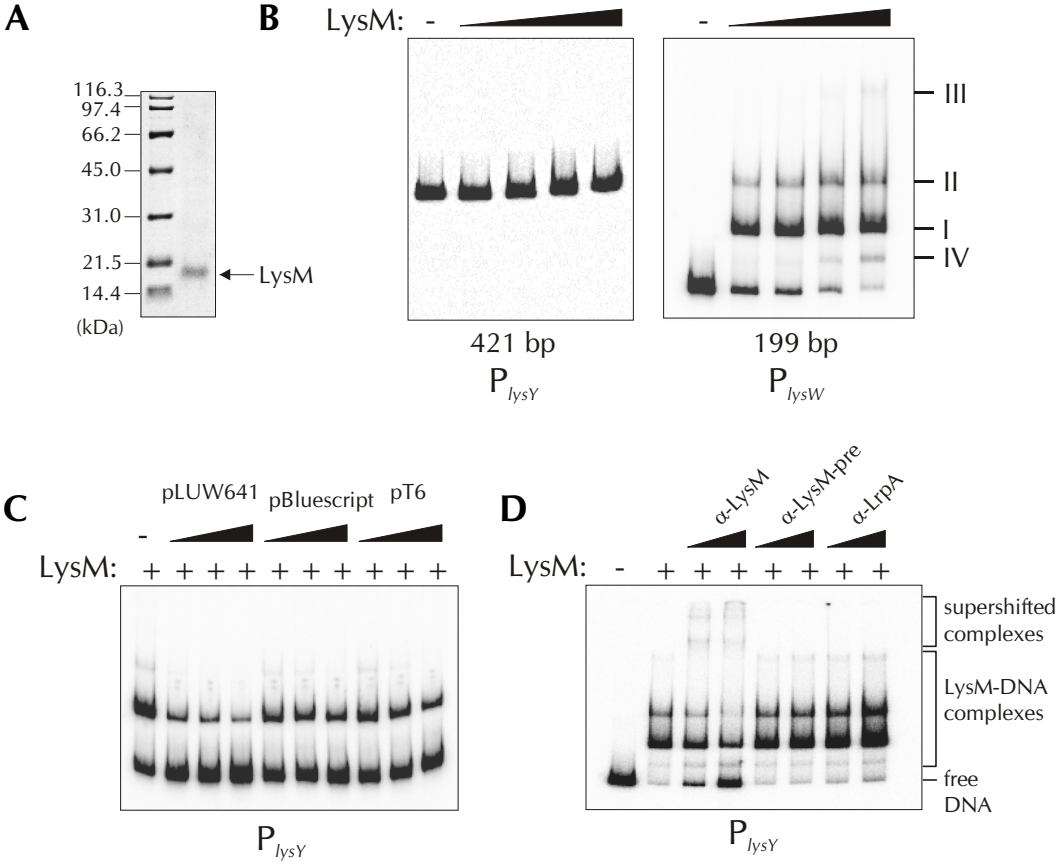



**Figure 3.** (**A**), SDS-PAGE analysis of purified LysM. (**B**), electrophoretic mobility shift assay (EMSA) showing that LysM binds to the *lysW* promoter ($P_{lysW}$, right), but not to the *lysY* promoter ($P_{lysY}$, left). Four distinct LysM-DNA complexes are formed upon addition of increasing amounts of LysM, as indicated (I to IV). 0, 30, 60, 200, and 600 ng of purified LysM was added, respectively. (**C**), competition assay showing that LysM-$P_{lysW}$ interaction is specific. Binding reactions were performed in the presence 5 ng of purified LysM, and 0.25, 0.75, or 2.5 µg of pBluescriptII SK+ plasmid DNA with different inserts. pLUW641 contains $P_{lysW}$ DNA, whereas pT6 contains the *Sulfolobus* SSV1 viral T6 promoter (45). (**D**), EMSA with $P_{lysW}$ DNA, LysM, and an antiserum raised against LysM (α-LysM). In the presence of α-LysM, LysM-DNA complexes either disappear or become supershifted by the LysM antibodies, as indicated (right). α-LrpA, antiserum raised against *P.furiosus* LrpA (9). 100 ng of LysM, and 200 or 400 ng of antiserum was used.

A 421-bp DNA fragment containing the *lysY* promoter ($P_{lysY}$) overlapping the BRE and TATA sequences was used. However, no binding of LysM to this fragment was observed (Fig. 3B). To assay LysM binding to the *lysW* promoter ($P_{lysW}$), a 199-bp DNA fragment containing the $P_{lysW}$ BRE and TATA sequences was used in the EMSA. The assay revealed that LysM binds to this DNA fragment, forming four protein-DNA complexes of distinct electrophoretic mobility (I-IV, Fig. 3B). Binding of LysM to $P_{lysW}$ appeared to be specific, since pLUW641, a plasmid containing $P_{lysW}$, competed for binding, whereas no competition occurred with the control plasmids pBluescript II SK+ (Stratagene) or pT6, containing the *Sulfolobus shibatae* virus SSV1 T6 promoter (45) (Fig. 3C). Using an antiserum raised against purified recombinant LysM, we verified whether the bands appearing upon addition of LysM represent LysM-DNA complexes. Upon addition of the antiserum, the LysM-DNA complexes either disappeared or supershifted, due to the binding of antibodies to the LysM protein (Fig. 3D). This effect was not observed using the LysM pre-immune serum or an antiserum against *P.furiosus* LrpA (9). To screen for additional LysM binding sites outside the two tested fragments, 1-kb DNA fragments overlapping the *lysY* and *lysW* promoters were digested with several restriction enzymes to obtain fragments between 50 and 500 bp. These fragments were subsequently used in an EMSA with LysM, revealing that only a restriction fragment overlapping the 199-bp $P_{lysW}$ fragment shifted (data not shown). We therefore conclude that LysM binds to a single site at $P_{lysW}$, but not to $P_{lysY}$. Binding experiments were performed at room temperature, 48ºC, or 65ºC, but no difference in affinity was observed (not shown).

## Lysine specifically affects LysM-DNA binding

Since transcription from $P_{lysW}$ varies strongly in response to the presence of lysine in the growth medium (Fig. 2), we hypothesized that lysine acts as a ligand for LysM, which in turn regulates transcription from $P_{lysW}$. Using EMSAs we studied the effect of lysine and several other amino acids on LysM-DNA binding. Addition of lysine in the binding reaction decreased the affinity of LysM for $P_{lysW}$, but did not completely eliminate binding (Fig. 4A). We tested higher lysine concentrations for complete inhibition of DNA-binding, but we found that throughout the tested concentration range (0.2 to 30 mM) the inhibition is constant (Fig. 4B). Binding inhibition was specific for lysine, since addition of other amino acids had no effect on LysM-DNA binding (Fig. 4A). To rule out the possibility that lysine is an aspecific DNA-binding inhibitor for Lrp-like proteins, we tested whether it also affected binding of the previously characterized *P.furiosus* LrpA protein to its promoter (9). However, lysine had no effect of LrpA-DNA binding (Fig. 4C). To analyze the effect of

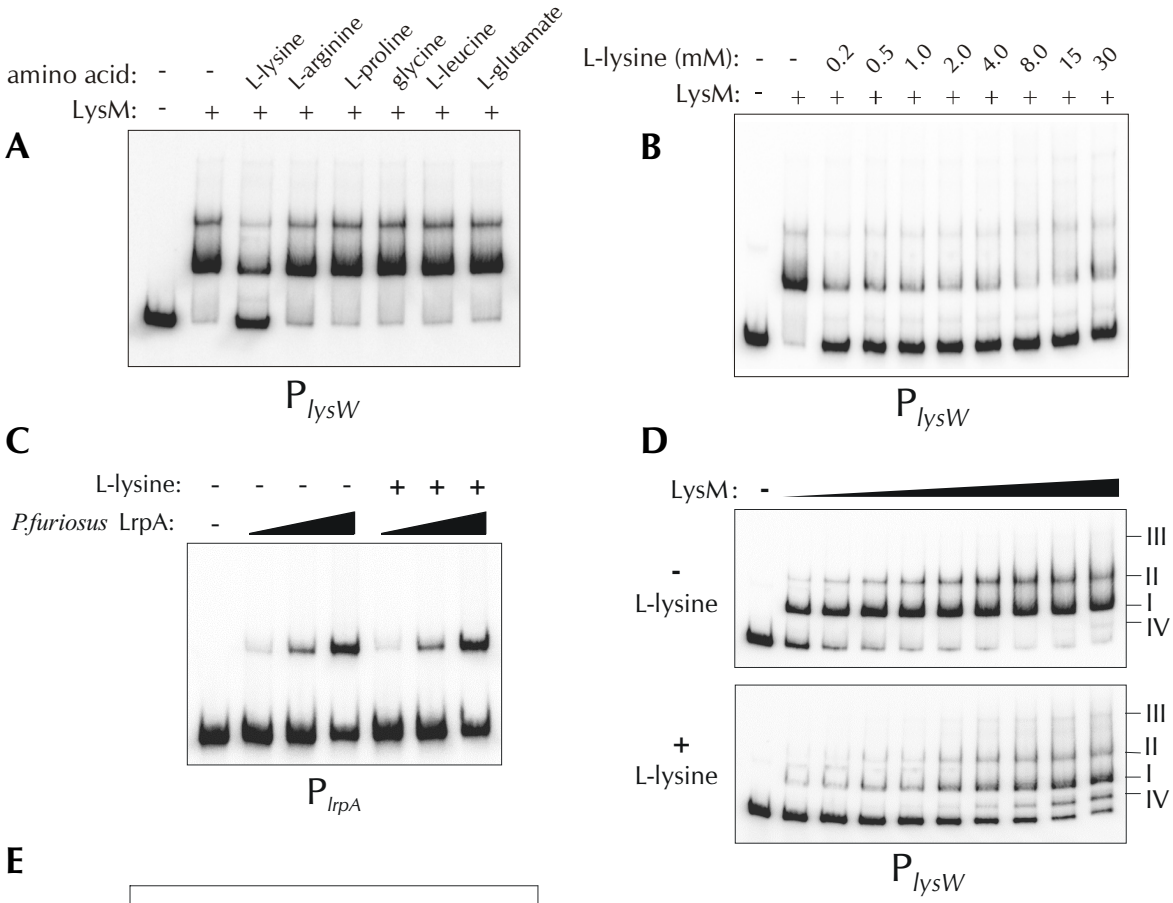



**Figure 4.** The effect of L-lysine on LysM-P*lysW* interaction. (**A**), several different amino acids were tested for their ability to affect LysM-P*lysW* interaction. 50 ng of LysM was used, and amino acids were added to a final concentration of 0.5 mM. (**B**), different concentrations of L-lysine were tested as indicated on top, 50 ng of LysM was used in the binding reaction. (**C**), L-lysine does not affect binding of *P.furiosus* LrpA to *lrpA* promoter DNA (9). 0, 50, 150,or 500 ng of purified LrpA was added, and L-lysine was added to a final concentration of 0.5 mM. (**D**), increasing amounts of LysM were used in an EMSA with P*lysW* DNA, in the presence or absence of 0.5 mM L-lysine. 0, 2.5, 5, 10, 20, 40, 80, 150, 300, or 600 ng of purified LysM was used in the binding reactions, respectively. (**E**), quantification of Fig. 4D. The percentage of radiolabeled DNA present in either of the LysM-DNA complexes was plotted against the amount of LysM added to the binding reactions.

lysine in more detail, we performed EMSAs with gradually increasing concentrations of LysM, in the presence or absence of lysine (Fig. 4D and E). Two major effects of lysine were apparent. First, lysine decreased the overall LysM-DNA binding affinity. Quantification of the results obtained in Fig. 4D revealed that the dissociation constant ($K_d$), defined as the LysM concentration at which 50% of the DNA is in complex with

LysM, is about 10 nM in the absence of lysine and 330 nM in the presence of lysine, reflecting a 33-fold decrease in DNA-binding affinity. Second, lysine changed the relative abundance of individual complexes. For example, complex IV is almost absent without lysine, whereas it is more abundant when lysine is present. This possibly reflects a structural change in one of the LysM-DNA complexes (e.g., increased compaction), or the formation of a novel complex, dependent on the presence of lysine. For *E.coli* Lrp it has been shown recently that its ligand, leucine, induces dissociation of the Lrp hexadecameric form to the octameric form (12), which may alter the affinity for sites within its different target operons. Using chemical cross-linking of LysM we analyzed its oligomeric state. LysM was incubated with or without lysine and cross-linked using dimethyl suberimidate (DMSI) (14). LysM cross-linked as a protein with a maximum molecular mass of about 66 kDa (not shown), which is indicative of a LysM tetramer. The addition of lysine had no apparent effect on cross-linking, suggesting that LysM multimerization is unaffected by lysine. However, this interpretation should be taken with care, since DMSI specifically cross-links the amino groups of lysine residues, and the addition of free lysine could therefore quench or interfere with the cross-linking reaction.

## LysM binds directly upstream of the BRE of $P_{lysW}$

Using DNaseI footprinting, we mapped the LysM binding site at the *lysW* promoter. We found that LysM protects a region of at least 15 bp directly upstream of the BRE of $P_{lysW}$ (Fig. 5A). Furthermore, bands representing sites hypersensitive to DNaseI cleavage appear outside of the LysM footprint. These sites are indicative of secondary structure changes of DNA, induced by LysM. Addition of lysine had no obvious effect on the footprint pattern, although some hypersensitive sites appeared to be slightly less abundant when lysine was present.

To confirm that the LysM binding site is indeed the sequence protected from DNaseI cleavage directly upstream of the $P_{lysW}$ BRE, an EMSA was performed with LysM and a 24-bp synthetic DNA fragment containing this LysM binding site (Fig. 5C). LysM bound to this fragment, and as with the larger fragment used for Fig. 4AB, lysine decreased the affinity of LysM for this fragment (Fig. 5B). In contrast to LysM binding to a larger fragment, which gives rise to the formation of four distinct LysM-DNA complexes, only two LysM-DNA complexes were formed (I, II). In addition, the overall affinity of LysM for this DNA is lower that for a larger fragment (compare Fig. 4 and 5B). Most likely, LysM specifically recognizes and binds the 24-bp sequence directly upstream of $P_{lysW}$, while

interactions between LysM and the DNA flanking this sequence may contribute to stronger binding and the formation of additional LysM-DNA complexes.

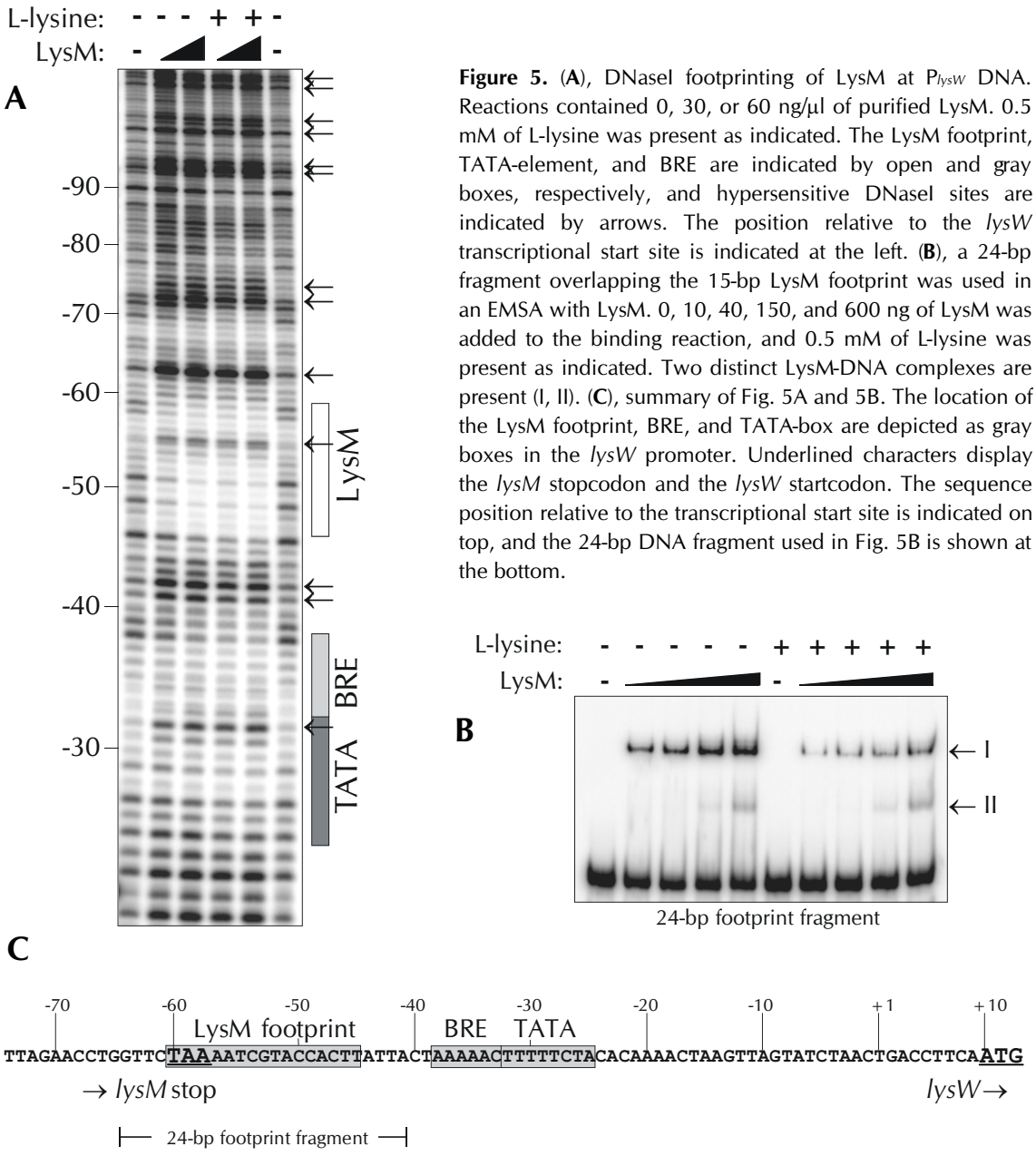



**Figure 5.** (**A**), DNaseI footprinting of LysM at $P_{lysW}$ DNA. Reactions contained 0, 30, or 60 ng/µl of purified LysM. 0.5 mM of L-lysine was present as indicated. The LysM footprint, TATA-element, and BRE are indicated by open and gray boxes, respectively, and hypersensitive DNaseI sites are indicated by arrows. The position relative to the *lysW* transcriptional start site is indicated at the left. (**B**), a 24-bp fragment overlapping the 15-bp LysM footprint was used in an EMSA with LysM. 0, 10, 40, 150, and 600 ng of LysM was added to the binding reaction, and 0.5 mM of L-lysine was present as indicated. Two distinct LysM-DNA complexes are present (I, II). (**C**), summary of Fig. 5A and 5B. The location of the LysM footprint, BRE, and TATA-box are depicted as gray boxes in the *lysW* promoter. Underlined characters display the *lysM* stopcodon and the *lysW* startcodon. The sequence position relative to the transcriptional start site is indicated on top, and the 24-bp DNA fragment used in Fig. 5B is shown at the bottom.

### *In vitro* transcription from $P_{lysY}$ and $P_{lysW}$

To study the effect of LysM on transcription of the *lys* cluster, we performed *in vitro* transcription experiments with *Sulfolobus* TBP, TFB, and RNAP (45). Transcription from the previously characterized T6 control promoter (45) was efficient, and a transcript of the

expected size was obtained (Fig. 6). Using the *lysY* promoter as a template also resulted in a transcript of the expected length, but transcription was less efficient compared to T6. Presumably, $P_{lysY}$ is an intrinsically weak promoter, and we previously showed that this promoter (referred to as the *argC* promoter) could be stimulated by the transcription factor TFE, which is dispensable for basal transcription *in vitro*, but stimulates transcription from sub-optimal promoters (2). In accordance with data obtained from EMSA experiments, addition of LysM had no effect on transcription from $P_{lysY}$, either in the presence or absence of lysine (not shown). Using $P_{lysW}$ in *in vitro* transcription experiments we could not detect any transcription, and addition of TFE or LysM, either with or without lysine, had no effect (not shown). Transcription reactions were therefore performed with cell extracts from *S.solfataricus* grown in the absence of lysine (where transcription from $P_{lysW}$ *in vivo* is maximal, see Fig. 2A), either in the presence (not shown) or absence (Fig. 6) of purified TBP, TFB, and RNAP. However, no transcription from $P_{lysW}$ could be detected.

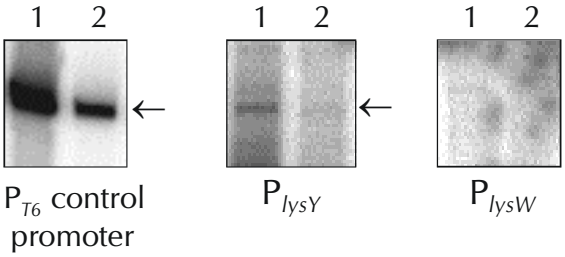



**Figure 6.** *In vitro* transcription using $P_{lysY}$ and $P_{lysW}$. 1, a reconstituted system with *Sulfolobus* TBP, TFB, and RNAP; 2, *in vitro* transcription system using only crude extracts from *Sulfolobus* grown on defined medium lacking lysine. Lower intensity of produced transcripts in these reactions is due to lower specific activity of transcription factors when compared to the purified system. Arrows indicate transcripts of expected size.

## $P_{lysY}$ and $P_{lysW}$ are bona fide promoters *in vivo*

Because $P_{lysW}$ had no detectable activity under the tested conditions, it had to be verified whether $P_{lysW}$ is a true promoter *in vivo*. In an alternative model for regulation of the *lysWXJK* transcript, transcription could be initiated from the (constitutive) *lysY* upstream promoter, and read-through into *lysWXJK* would be allowed only when lysine is absent. In this hypothetical model LysM could act as a transcriptional roadblock which, in response to lysine, permits RNA polymerase to read through or terminate transcription at or near the LysM binding site. In the absence of lysine this would give rise to a transcript of about 5.5 kb containing *lysYZMWXJK*. Since no mRNA of this size was detected by Northern blotting (Fig. 2A), a subsequent RNase processing event would be responsible for cleavage of *lysYZMWXJK* into *lysYZM* and *lysWXJK*, which 5' terminus we have mapped in primer extension and RNase protection experiments (Fig. 2D and E). To investigate the validity of this hypothetical model it was crucial to determine whether the 5' terminus of the *lysWXJK* is initiated or processed. Therefore we performed ligation-mediated RT-PCR as described

by Bensing et al. (7). Briefly, total RNA is either treated or not treated with tobacco acid phosphatase (TAP), which converts 5' triphosphate termini, indicative of transcription initiation, into 5' monophosphate termini, indicative of processing. Only when a 5'monophosphate is present, an RNA adapter can subsequently be ligated to these termini. In an RT-PCR with adapter-specific and gene-specific primers, only mRNAs that had 5' monophosphates can thus be amplified, and in this way it is possible to discriminate between initiated and processed 5' mRNA termini. With this technique we analyzed 5' termini of *lysY*, *lysW*, as well as the *16S* gene. The latter has been shown to be processed in *S.solfataricus* (13, 46). For all three tested 5' ends we obtained PCR products of expected length, based on our primer extension and RNase protection data. As expected, *16S* control RNA could be PCR-amplified both from TAP-untreated as well as for TAP-treated samples (Fig. 7), which indicated that its 5' end is the result of RNA processing. *lysY* and *lysW* RNA could only be PCR-amplified from TAP-treated samples, which demonstrated that both *lysY* and *lysW* promoters have initiated 5' ends (Fig. 7). We concluded therefore that $P_{lysY}$ and $P_{lysW}$ are bona fide promoters *in vivo*, however, under the used *in vitro* transcription reaction conditions transcription does not take place.

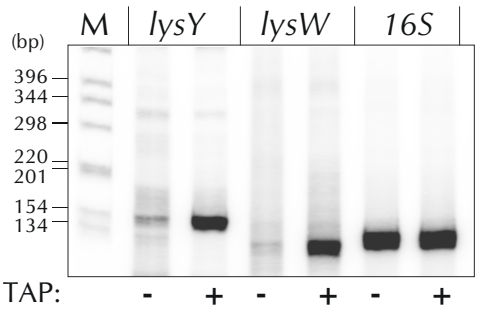



**Figure 7.** Discrimination between transcription initiation or processing at the 5' termini of *lysYZM* and *lysWXJK* transcripts, using tobacco acid phosphatase (TAP) and ligation-mediated RT-PCR. Initiated 5' termini can only be PCR-amplified when their 5'-triphosphates are converted to 5'-monophosphates by TAP, thus enabling the ligation of an RNA linker. Processed 5' termini contain 5'-monophosphates and an RNA linker can be ligated independently of TAP treatment. DNA fragments were RT-PCR-amplified using a linker-specific primer and a gene-specific primer. M, DNA marker. Expected sizes of PCR products are 146 bp for *lysY*, 124 bp for *lysW*, and 132 bp for *16S*, a positive control for the detection of a processed 5' terminus.

## *DISCUSSION*

Genomic sequence data has revealed that Lrp-like proteins are ubiquitously present in bacteria and archaea. Although Lrp from *E.coli* is the archetype Lrp-like protein, more than a dozen bacterial Lrp-like proteins have been studied either *in vitro* or using genetic approaches, and it is clear that these proteins are all involved in the regulation of amino acid metabolism, and that amino acids serve as ligands for Lrp-like proteins (10). Efficient

genetic tools to study gene regulation (like recombination or gene-disruption) are not yet available for hyperthermophilic archaea, and we have used a bioinformatics approach in combination with *in vivo* and *in vitro* analyses to identify and study the archaeal Lrp-like protein LysM from *S.solfataricus*, the gene of which is present in the *lys* gene cluster. This allowed us for the first time to study the role and function of an archaeal Lrp-like protein in relation to its physiological target genes.

Although most of the *S.solfataricus lys* genes have homology to classical arginine biosynthesis genes, the repressive effect of lysine strongly suggests that the cluster is involved in lysine biosynthesis. This is in agreement with the lysine auxotrophic phenotype of gene-disruption mutants in the analogous *T.thermophilus lys* gene cluster, where lysine is synthesized via a novel biosynthetic route involving α-aminoadipic acid (AAA) rather than diaminopimelic acid (DAA), the typical bacterial precursor (26, 27, 42). Moreover, some essential enzymes of the DAA pathway appear not to be encoded by the *S.solfataricus* genome (e.g. dihydrodipicolinate reductase, diaminopimelate epimerase, and diaminopimelate decarboxylase). Hence, it is expected that like *T.thermophilus*, *S.solfataricus* uses the AAA pathway for lysine biosynthesis. However, our data does not exclude the possibility that the encoded enzymes are also functional in arginine biosynthesis. Possibly, the *lys* gene cluster encodes bi-functional enzymes, involved in both lysine and arginine biosynthesis, as was proposed for the *lys* gene cluster of *P.horikoshii* (42). Dual activity has been measured for LysJ and LysK of *T.thermophilus*, homologues of *S.solfataricus* LysJ and LysK, respectively, suggesting that it could indeed be active in arginine as well as lysine biosynthesis (38, 39). If the *lys* cluster indeed encodes lysine as well as arginine biosynthesis enzymes, arginine biosynthesis is thus regulated only by the presence of lysine, and not arginine. The following observation supports this hypothesis. When *S.solfataricus* was grown in the presence of only lysine, we found that the generation time was more than two-fold longer than during growth in the presence of only arginine, both arginine and lysine, or in the absence or presence of all amino acids. This suggests that the expression of arginine biosynthesis enzymes is down-regulated while their activities are necessary under this particular condition. These enzymes could very well be encoded by the *lys* gene cluster, since part of it is downregulated by lysine. Moreover, whereas the *S.solfataricus* genome encodes additional paralogues of LysY (ArgC), LysJ (ArgD), and LysK (ArgE), there is only a singly copy of LysZ (ArgB), which means that LysZ encoded by the *lys* gene cluster is essential for arginine biosynthesis. In this respect it is interesting to mention earlier work by Van de Casteele *et al.* (53), who measured activities of arginine biosynthesis enzymes in *S.solfataricus* cell extracts. It was shown that arginine in the medium did not reduce any of

the measured specific enzyme activities, except for that of *N*-acetylglutamate kinase, which was down-regulated three-fold. The *lysZ* gene is the only potential candidate gene in *S.solfataricus* encoding this enzyme, but according our results there is no regulation of this gene at the level of transcription. Since feedback inhibition of *N*-acetylglutamate kinase by arginine was excluded, it has to be concluded that the expression of LysZ is regulated post-transcriptionally.

Why is the *lys* gene cluster of *S.solfataricus* organized in a constitutive (*lysYZM*) and a regulated part (*lysWXJK*)? Possibly, down-regulation of the *lysYZM* is not permitted because this would abolish regulation of the *lysWXJK* genes through LysM, the gene of which is co-transcribed with *lysZY*. Our data does not exclude the possibility that LysM has additional targets in the genome of *S.solfataricus*, but if this were the case, down-regulation of *lysM* could result in an even more serious loss of regulatory capacity in *S.solfataricus*. Alternatively, one of the enzyme activities encoded by *lysYZM* could be indispensable for growth under the conditions tested. The genomic organization of *lysM* in the *lys* cluster is identical in other *Sulfolobus* species, and functionally comparable to that of *A.pernix*, where the *lysYZM* cluster is inverted (Fig. 1A), most likely allowing a similar mode of regulation.

Although the role of *lysW* and *lysX* has not been demonstrated experimentally, we speculate that they are specific for the prokaryotic AAA lysine biosynthesis route proposed by Nishida *et al.* (42), since they are clustered within the *lys* clusters of *T.thermophilus, S.acidocaldarius, S.tokodai*, *A.pernix*, *Pyrococcus* species, and *T.acidarmanus*, and because the classical arginine or lysine pathways do not involve such genes. LysX is 24% identical to RimK of *E.coli*, which was shown to be a post-translational modification enzyme, catalyzing the coupling of four glutamate residues to the C-terminus of the S6 ribosomal protein (21). However, we found that in several prokaryotic genomes *rimK*-like genes are clustered with amino acid biosynthesis genes (not shown), suggesting that its catalytic activity here is not utilized for a post-translational modification, but rather in amino acid biosynthesis. Interestingly, Galperin *et al.* (18) demonstrated that RimK belongs to a superfamily of enzymes with a so-called 'ATP-grasp' fold. This family includes enzymes like D-alanine-D-alanine ligase, glutathione synthetase, biotin carboxylase, and carbamoyl phosphate synthetase. All these enzymes possess ATP-dependent carboxylate-amine ligase activity, i.e. the capacity to form a peptide bond. In the proposed AAA-dependent lysine biosynthesis route described by Nishida *et al.* (42), LysX was predicted to catalyze a similar reaction, namely connecting the amino group of AAA to the carboxyl group of a yet unidentified molecule. By doing so, LysX catalyzes a reaction functionally analogous to that of ArgA

(*N*-acteylglutamate synthase) in the classical arginine biosynthesis pathway, the gene of which is absent in all *lys* clusters shown in Figure 1A.

The small protein encoded by *lysW* has no homologues in the database, apart from *lysW*-genes found in the *lys* clusters depicted in Figure 1A. A PSI-BLAST analysis (1) of LysW showed that this small protein is homologous to the N-terminal domain of archaeal TFB transcription factors (not shown). After four iterations an expect value of $\geq 10^{-11}$ was obtained with TFB N-terminal domains from several archaea. This domain consists of a zinc-ribbon (56), and the two CPxCG 'zinc knuckle' motifs that bind the zinc atom are well conserved in LysW. The zinc-ribbon domain of *Sulfolobus* TFB has been shown to be involved in the recruitment of RNA polymerase (RNAP) (5), through interaction with the RpoK subunit of RNAP (35). Generally, zinc-ribbon domains mediate protein-protein or protein-DNA interactions and can be found for instance in several eukaryal transcription factors. LysW may therefore interact with one of the encoded enzymes of the *lys* gene cluster, perhaps acting as a regulatory subunit for the respective enzymatic activity, or it could play a regulatory role in *lys* transcription. However, LysW is also encoded by the *lys* gene cluster of *T.thermophilus*, which possesses a bacterial basal transcription machinery, and it is questionable whether involvement in transcription here is possible. It should be noted though that the RpoK subunit, with which the TFB zinc-ribbon interacts, is a conserved subunit in RNAPs of eukarya (RPB6) and bacteria (ω). Unfortunately, our attempts to produce the LysW protein in *E.coli* were unsuccessful.

The data presented in this study strongly suggests that LysM acts as a transcriptional activator for $P_{lysW}$. We found that transcription from $P_{lysW}$ is maximal when lysine is absent, and under these conditions the affinity of LysM for binding this promoter is the highest. Conversely, in the presence of lysine transcription is lowest, and the binding affinity of LysM is decreased. It is most likely that LysM occupies $P_{lysW}$ preferably when lysine is absent, thereby somehow activating this promoter. Although lysine reduces rather than eliminates LysM-DNA binding, this reduction is expected to be physiologically important. As shown in Figure 4D and 4E, the effect of lysine is maximal at a low LysM concentration, presumably a relevant cellular condition (using Western blotting analysis we roughly estimated the abundance of LysM to be about 0.01% of total soluble protein, not shown). This reduction rather than elimination of binding has also been observed for other (bacterial) Lrp-like proteins (17, 20, 33, 48, 55), and may therefore be a general feature of Lrp-like proteins. Using *in vivo* formaldehyde cross-linking followed by immunoprecipitation of cross-linked LysM-DNA complexes, we have attempted to relate results from *in vitro* binding studies with *in vivo* LysM-promoter occupation, but

unfortunately the results of these experiments were irreproducible. Nevertheless, to a certain extent our study is comparable to the regulation of *E.coli ilvIH* by Lrp. For example, Lrp activates *ilvIH* transcription, and this activation is decreased when leucine is present in the medium (44). In accordance, the Lrp-*ilvIH* affinity *in vitro* is reduced but not eliminated by leucine (47). For *ilvIH* this reduction of binding *in vitro* could be related to an *in vivo* decrease in promoter occupancy using *in vivo* footprinting experiments (36), and we anticipate that this *in vitro-in vivo* relationship can also be made for LysM.

Our DNaseI footprint data supports the possibility that LysM is an activator, since it showed that LysM protects the bases -46 to -59 relative to the transcriptional start site, whereas the TBP-TFB-RNAP pre-initiation complex has previously been shown to protect the bases -43 to +8 at *Sulfolobus* viral T6 promoter (5). Hence, LysM binding is not expected to interfere with or occupy the target sites for TBP, TFB, or RNAP, but rather binds upstream of these proteins, as is usually the case for activators. In many cases, these activators recruit components or stabilize binding of the pre-initiation complex (PIC) through direct contacts. However, such direct contacts have not yet been shown for Lrp-like proteins. Alternatively, activation could involve promoter remodeling in which the secondary structure of the promoter DNA is changed, for instance by bending the DNA in a certain angle. This altered DNA structure could subsequently be recognized more efficiently by one of the components of the pre-initiation complex (PIC). In this case, activation of transcription would be independent of direct contacts between the activator and the PIC components. As shown in Figure 5A, binding of LysM induces several DNaseI hypersensitive sites, indicative of DNA secondary structure changes. Interestingly, one of the hypersensitive sites is located between the BRE and TATA-box, representing a structural alteration that is potentially able to alter the interaction properties of TBP/TFB with the DNA. In addition, the intensity of some of the LysM-induced hypersensitive sites is somewhat changed in the presence of lysine. However, the magnitude of this effect is less obvious than the effect of lysine observed in EMSAs (Fig. 4). We have tested the ability of LysM to bend its target DNA by using the pBEND2 system described by Kim *et al.* (25). For this purpose we cloned the 24-bp fragment used for Figure 5B and 5C into pBEND2, and used it as described (25), but no LysM-induced bending was observed, either in the presence of absence of lysine. It cannot be ruled out, however, that low-affinity binding to sequences outside the chosen sequence fragment contribute to the LysM-DNA interaction and bending.

Our *in vitro* transcription experiments showed that compared to the T6 control promoter (45), transcription from $P_{lysY}$ and $P_{lysW}$ is very weak. We showed previously that

$P_{lysY}$ could be stimulated by the addition of TFE (2) (referred to as $P_{argC}$), but this was not possible for $P_{lysW}$. Apparently, both promoters are intrinsically weak promoters, which is in agreement with the low homology to the *Sulfolobus* consensus promoter sequence. It is therefore possible that binding of TBP and/or TFB might be impaired at $P_{lysW}$. In agreement with this, in an EMSA we could not observe any interaction between $P_{lysW}$ DNA and (combinations of) these transcription factors, and addition of LysM or lysine had no effect (not shown). Altogether, our results thus suggest that the intrinsic activity of $P_{lysW}$ promoters is very low. In contrast, *lysWXJK* mRNA could easily be detected in Northern blotting and primer extension experiments, suggesting efficient transcription *in vivo*. Since we have proven that both $P_{lysY}$ and $P_{lysW}$ are true promoters *in vivo*, we suggest that at least for $P_{lysW}$, additional factors like co-regulators may be required for efficient transcription. Thus, under our experimental conditions LysM is not able to activate transcription, but in the presence of such an unidentified factor transcription may take place. In comparison, some *E.coli* promoters that belong to the Lrp regulon require an additional DNA-binding protein, e.g. integration host factor (IHF) (43), histone-like protein H-NS (30), or catabolite activator protein (CAP) (37, 54). To identify such proteins in *S.solfataricus* we have taken several approaches. First, in our *in vitro* transcription experiments we have added cell extracts of *S.solfataricus* grown in the absence of lysine, but no stimulation of transcription *in vitro* was observed. Second, we have performed pull-down experiments in which glutathione-S-transferase (GST) was fused to LysM and immobilized on glutathione agarose beads. A *S.solfataricus* cell extract was subsequently screened for proteins interacting with LysM. We found that a single protein interacted with LysM, but this protein was identified as the LysM protein itself, most likely being the result of multimerization of GST-LysM and wild-type LysM during the experiment (data not shown).

The observation that LysM is conserved in the *lys* clusters of three *Sulfolobus* species and in *A.pernix* suggests that it plays a similar role in these organisms. We have therefore compared the sequence of their putative *lysW* promoters to identify a possible consensus LysM binding site. Indeed, a conserved GGTTC inverted repeat element is present, as shown in Figure 8. For the putative *lysW* promoters of *Sulfolobus* species, the position of the presumptive LysM binding site is very similar (overlapping the *lysM* stop codon), whereas in *A.pernix*, where the *lys* gene cluster is organized in a somewhat different way (see Fig. 1A), the putative LysM site is centered between the *lysY* and *lysW* genes. It is remarkable that this LysM binding site is conserved and highly palindromic, since this is usually not the case for naturally occurring binding sites of Lrp-like proteins (9, 11, 34). We have derived a consensus LysM-site from the alignment given in Fig. 8, and we used this sequence to

screen the *S.solfataricus* and *A.pernix* genomes for LysM sites, but no additional LysM binding sites could be identified.

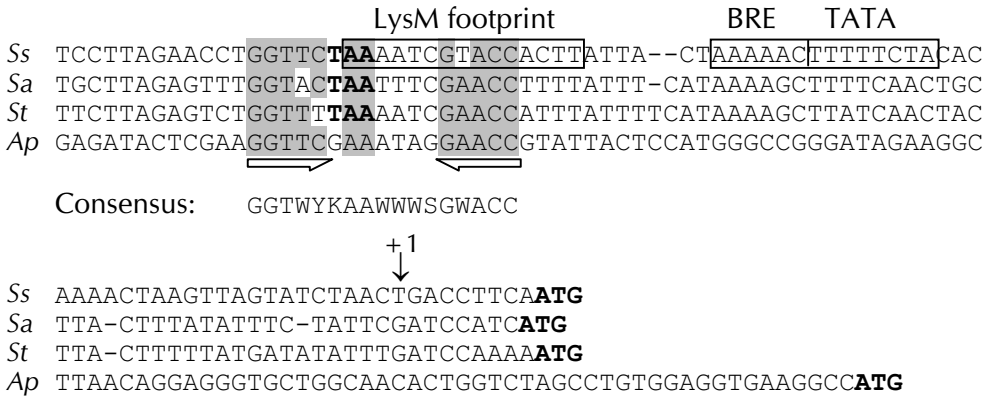



**Figure 8.** Alignment of (putative) *lysW* promoters of *S.solfataricus (Ss)*, *S.acidocaldarius (Sa)*, *S.tokodai (St)*, and *A.pernix (Ap)*. Gray shading indicates the conserved sequence of the LysM binding site, and open arrows indicate inverted repeat elements. Bold characters display stop codons and start codons of the *lysM* and *lysW* genes. The *S.solfataricus* LysM footprint as well as the $P_{lysW}$ BRE and TATA elements are boxed, and the transcriptional start site (+1) is indicated (vertical arrow). A consensus LysM site was derived from the alignment, and is given below (W = A/T; Y = C/T; K = G/T; S = G/C).

Several archaeal Lrp-like proteins have been characterized recently (4, 9, 15, 40), but LysM is the first for which a clear physiological role has been demonstrated. Unlike previously studied archaeal Lrp-like proteins, *lysM* is expressed constitutively and not negatively autoregulated. Moreover, LysM strongly resembles bacterial Lrp-like proteins, and appears to be a specific rather than a global regulator, since it is clustered with its target genes. However, we cannot exclude the possibility that LysM has additional targets in the *S.solfataricus* genome, and experiments are necessary to confirm this. Furthermore, all bacterial Lrp-like proteins characterized to date act as transcriptional repressors or activators involved in the regulation of amino acid metabolism, and all ligands identified so far were found to be amino acids. In analogy, the previously characterized archaeal Lrp's are repressors, whereas our data is most compatible with LysM being an activator. Thus, Lrp-like proteins appear to be functionally equivalent in the bacterial and archaeal domains, despite the fundamental differences in transcriptional machineries.

In conclusion, we have studied the role of the *S.solfataricus* Lrp-like protein LysM in the regulation of the *S.solfataricus lysYZMWXJK* gene cluster, which is involved in lysine and possibly arginine biosynthesis. Transcription of this cluster arises from two promoters, $P_{lysY}$ and $P_{lysW}$, and addition of lysine (but not arginine) represses the internal $P_{lysW}$ activity eight-fold without affecting $P_{lysY}$. LysM binds to a site directly upstream from the BRE of $P_{lysW}$,

and since the affinity of LysM for this binding site *in vitro* is highest in the absence of lysine, it is most likely that LysM acts as an activator for *lysWXJK* transcription. The fact that we could not confirm this in an *in vitro* transcription system does suggest that activation by LysM requires one or more (yet unidentified) additional factors, as is the case for transcriptional activation of some promoters by the homologous *E.coli* Lrp. Future research will be focused on the identification of these cofactors using the yeast two-hybrid system. Furthermore, transcription and regulation of the $P_{lysW}$ promoter *in vitro* could be further analyzed in conjunction with the recently characterized chromatin-associated protein Alba, which has been shown to possess transcription regulatory potential.

## ACKNOWLEDGEMENTS

This research was supported by the Council for Chemical Sciences (CW) of the Netherlands Organization for Scientific research (NWO) grant 700-35-101. We thank Dr. Q. She (Copenhagen) for providing the sequence of the *S.acidocaldarius lys* gene cluster.

## EXPERIMENTAL PROCEDURES

### Growth of *Sulfolobus solfataricus*

*Sulfolobus solfataricus* P2 was grown in defined medium containing 3.1 g/l $KH_2PO_4$, 2.5 g/l $(NH_4)_2SO_4$, 0.2 g/l $MgSO_4 \cdot 7H_2O$, 0.25 g/l $CaCl_2 \cdot 2H_2O$, 1.8 mg/l $MnCl_2 \cdot 4H_2O$, 4.5 mg/l $Na_2B_4O_7 \cdot 10H_2O$, 0.22 mg/l $ZnSO_4 \cdot 7H_2O$, 0.06 mg/l $CuCl_2$, 0.03 mg/l $Na_2MoO_4 \cdot 2H_2O$, 0.03 mg/l $VOSO_4 \cdot 2H_2O$, 0.01mg/l $CoCl_2$, supplemented with 2 g/l sucrose, 0.02 g/l $FeCl_3$ and vitamins, adjusted to pH 3.0 with $H_2SO_4$. Amino acids, if present, were added to a final concentration of 1 mM.

### Analysis of *S.solfataricus* RNA

Total RNA was isolated from *S.solfataricus* using the RNeasy method (QIAgen). 35 ml of mid-log phase culture ($OD_{600}$ of 0.4) was washed in 1 ml of medium and resuspended in 100 µl of TE. 5 µl of 10% Triton X-100 was added, and further purification was done according the manufacturer's prescriptions, except that genomic DNA was sheared through a 0.45 mm needle before the sample was applied unto the column. Columns were eluted twice with 50 µl of water.

For Northern analysis 15 μg of total RNA was separated on a 1.3% formaldehyde agarose gel and blotted to a Hybond $N^+$ membrane as described by Sambrook et al. (49). Radiolabeled DNA probes were hybridized using Ultrahyb™ solution (Ambion), according the manufacturer's prescriptions.

For primer extension analysis 10 μg of total RNA and 2.5 ng of radiolabeled oligo BG815 or BG876 was resuspended in 2x AMV-RT buffer (Promega) in a final volume of 25 μl. Samples were heated to 70°C for 10 min., and slowly cooled to room temperature. $MgCl_2$, dNTPs, RNasin, and AMV-RT (Promega) were added to a concentration of 5 mM, 0.4 mM, 0.8 U/μl, and 0.4 U/μl, respectively, in a final volume of 50 μl. The samples were incubated at 42°C for 30 min., extracted with phenol:chloroform, precipitated with ethanol, resuspended in formamide loading buffer, and analyzed on a 8% denaturing sequencing gel.

| Name | Sequence (5' → 3') |
|---|---|
| BG774 | CGCGCG*CCATGG*GTAATGCAAATATTGATGAAAG |
| BG775 | GCGCGC*GGATCC*TTAGAACCAGGTTCTAAGGATAATAG |
| BG876 | CTCTCCCCAATCCTCACCTATTTGC |
| BG815 | GTTATGACTGATATTTCTGTCTTCGG |
| BG816 | GCTGAAAAATATGACTTGCTTAACGG |
| BG876 | CTCTCCCCAATCCTCACCTATTTGC |
| BG877 | CCATTGAAGGTCAGTTAGATACTAAC |
| BG878 | GATTATGATATTCTAGTAATAGTTAGAGG |
| BG938 | GCGCGC*TCATGA*GGCACCACCACCACCACCACGCCATGGGTAATGCAAATATTGATGAAA GTG |
| BG1087 | GATTAGAAGCATTCAAGGTGTTGTG |
| BG1088 | TCTACAAGTTCTCCTGGTAACGC |
| BG1128 | AGCCGTCCGCCACGCTCCC |
| BG1133 | GAATAGCCCATGAGCCTTTCAG |
| BG1134 | AAATTCGGATGTATCAATGATACTGG |
| BG1198 | CTTTTATGGTAGGTATATAACCGG |
| BG1199 | TATCCTGAACCTCCAACTACGGC |

**Table I.** Oligonucleotides used in this study. Restriction sites are underlined and italic.

RNase protection was performed using RPA III™ and MAXIscript™ kits (Ambion), according the manufacturer's prescriptions. 10 μg of total *S.solfataricus* RNA was used, isolated according the method described above. For the generation of a labeled *lysW* antisense RNA probe, a 267-bp PCR fragment, amplified using the oligos BG814 and BG876 (see Table I), was cloned into an *Xcm*I-digested pBluescriptII SK(+)-derivative (8), yielding pLUW646. The orientation of the fragment was selected so that *in vitro* transcription with BamHI-digested pLUW646 and T3 RNA polymerase yielded a 367-nt *lysW* antisense RNA probe. For the generation of labeled RNA marker fragments, an unrelated 76-bp was cloned as described above, yielding pLUW647, which was digested either with *Eco*RI, *Eco*RV, *Cla*I, *Xho*I, or *Apa*I, and used in an *in vitro* transcription reaction

with T7 RNA polymerase, yielding RNA marker fragments of 153, 161, 172, 186, and 199 nt, respectively. RNase protection samples were analyzed on a 8% denaturing sequencing gel.

## Production and purification of *Sulfolobus solfataricus* LysM

The gene encoding *lysM* was PCR amplified using primers BG774 and BG775 (see Table I). Underlined sequences indicate the restriction sites *Nco*I and *Bam*HI. The resulting PCR fragment was digested with *Nco*I and *Bam*HI and cloned into the T7 expression vector pET24d (51) (Novagen, Inc.), resulting in the construct pLUW632. This construct was transformed into *E.coli* BL21(DE3) (Novagen, Inc.). A single colony was used to inoculate 5 ml of LB medium with 50 μg/ml kanamycin, and the culture was incubated in a rotary shaker at 37 °C until log-phase growth was observed. Subsequently, the culture was used to inoculate 1 liter of identical medium, and incubation was continued until an $OD_{600}$ of 0.5 was reached. Expression was induced by the addition of 0.5 mM IPTG, and incubation was continued for 2.5 hours. After expression the cells were harvested, washed in 20 mM Tris-HCl buffer pH 8.0. Cells from 250 ml of culture were resuspended in 15 ml of 20 mM Tris pH 8.0, and lysed by a single passage through a French pressure cell at 1000 psi. The lysate was centrifuged at 20,000 RPM for 20 min. and loaded on an 25-ml Q-sepharose column (Amersham Pharmacia Biotech) that had been equilibrated with 20 mM Tris pH 8.0. The flowthrough containing LysM was collected and subjected to a heat incubation at 80ºC for 30 min., and subsequently centrifuged for 20 min. at 20,000 RPM. The supernatant contained the purified LysM. $His_6$-LysM was used for the production of a rabbit-antiserum. For this purpose we cloned and expressed $His_6$-LysM as described above, using the oligo BG938 instead of BG775 (see Table I). The underlined sequence indicates a *Bsp*HI restriction site. $His_6$-LysM precipitated spontaneously from a cell extract after overnight storage at 4ºC, and we used this precipitated material to purify $His_6$-LysM under denaturing conditions using 8 M urea and Ni-NTA spin columns (QIAgen), according the manufacturer's prescriptions. The purified $His_6$-LysM was dialyzed stepwise against 50 mM Tris pH 8.0. $His_6$-LysM was only used for immunization since the protein lost its DNA-binding activity, most likely the result of unsuccessful renaturation.

## EMSAs and DNaseI footprinting

DNA probes used for gel-mobility shift experiments were generated using PCR. The following primers were used: BG816 and BG815 for a 421-bp $P_{lysY}$ fragment; BG878 and BG877 for a 199-bp $P_{lysW}$ fragment; BG1087 and BG1088 for a 207-bp $P_{lysW}$ fragment (see Table 1). PCR products were end-labeled using T4 kinase and radioactive $\gamma^{32}$P-ATP

(Amersham Pharmacia Biotech), and purified from a 6% acrylamide gel as described (49). Binding reactions were performed in a total volume of 10 µl, containing 50 mM Tris pH 8.0, 1 mM DTT, 10% glycerol, and varying concentrations of purified LysM. Standard reactions contained 1 to 10 ng of $\gamma^{32}$P-ATP end-labeled DNA and 50 ng of poly(dI.dC).poly(dI.dC) as nonspecific competitor DNA (Amersham Pharmacia Biotech). L-lysine, if present, was added to a concentration of 0.5 mM. Reactions were incubated at room temperature or at 48ºC for at least 10 min. and separated on a non-denaturing 6% acrylamide gel, buffered in 1x TBE buffer. Gels were dried, exposed to phosphor screens and analyzed. Probes for DNaseI footprinting were generated using PCR with the oligos BG877 and BG878, where one of the two oligos was end-labeled using T4 kinase and radioactive $\gamma^{32}$P-ATP. Probes were purified from 6% acrylamide (49). Binding reactions were performed at 48°C in a volume of 50 µl containing 50 mM Tris pH 8.0, 25 mM $MgCl_2$, 75 mM KCl, 1 mM DTT, and 100 ng of poly(dI.dC).poly(dI.dC). L-lysine, if present, was added to a concentration of 0.5 mM. After 10 min. 1 µl of a 1/50 dilution (about 0.6 U) of RNase-free DNaseI (Roche) was added, and incubation was continued for 1 min. The reaction was stopped and the samples were purified using phenol:chloroform extraction and ethanol precipitation. After resuspension in formamide loading buffer the samples were analyzed on a 8% denaturing sequencing gel.

## Chemical cross-linking and Western blotting of LysM

Chemical cross-linking and Western blotting of LysM was performed as described previously (9).

## *In vitro* transcription

*In vitro* transcription reactions with a reconstituted system contained 20 ng of recombinant TBP, 20 ng of recombinant TFB, and 100 ng of RNAP purified from *Sulfolobus* and 100 ng of undigested plasmid pT6, pLUW637, or pLUW648 template DNA. In reactions with crude *S.solfataricus* extracts purified TBP, TFB, and RNAP were either omitted or replaced by various amounts (up to 10 µg) of extract from *S.solfataricus* cells that were grown in the absence of lysine. The templates used in the assays consisted of promoter DNA cloned into pBluescript SK (Stratagene) or pBluescript SK derivatives (8). pT6 contained 207 bp of the T6 promoter from the SSV1 virus (45), pLUW637 contained 421 bp (-313 to +108) of the *lysY* promoter (see above), pLUW648 contained 379 bp (-277 to +102) of the *lysW* promoter. Reactions were performed in a volume of 50 µl in 50 mM Tris, pH 8.0, 75 mM KCl, 25 mM $MgCl_2$,1 mM dithiothreitol, 200 µM NTPs for 20 min at 70 °C. Reactions were terminated by the addition of 250 µl of 10 mM Tris, pH 8.0, 10 mM

EDTA, 750 mM NaCl, 1% SDS containing ~1 ng of $^{32}$P-5'-endlabeled antisense primer and 10 μg of glycogen. Following extraction with phenol/chloroform, nucleic acids were recovered by ethanol precipitation and resuspended in 20 μl of 1x AMV reverse transcriptase buffer containing200 μM dNTPs. After addition of 5 units of AMV reverse transcriptase (Roche) and incubation at 42 °C for 30 min, 20 μl of 95%formamide with 0.05% bromophenol blue were added, the reaction boiled for 3 min, and 20 μl loaded on a 8% polyacrylamide gel containing 8 M urea.

## Discrimination between initiated and processed 5' mRNA termini

The method was modified from the previously described method by Bensing et al. (7). 25 μg of total RNA from S.solfataricus grown in defined medium without any amino acids, was purified as described above and treated with 20 U of RNase-free DNaseI (Ambion) at 37º for one hour, in final volume of 100 μl 1x DNaseI buffer (Ambion). DNA-free RNA was subsequently purified using RNeasy spin columns (QIAgen) according the manufacturer's prescriptions, and 10 μg of this RNA was used in subsequent reactions. Treatment of RNA with TAP (tobacco acid phosphatase), and ligation of 5' RNA adapters to the 5' terminal ends was done using the FirstChoice™ RLM-RACE Kit (Ambion), according the manufacturer's protocol. TAP-untreated control reactions were identical to TAP-treated reactions, except that TAP was replaced by nuclease-free water. 1 μl of adapter-ligated RNA was used for reverse transcription with 100 ng of a gene-specific oligo (BG1134 for lysY, BG876 for lysW, and BG1133 for 16S), and 5 μl of 10x RT buffer (Promega), in a final volume of 25 μl. Samples were heated to 70ºC for 10 min. and slowly cooled to room temperature. $MgCl_2$, dNTPs, Rnasin, and AMV-RT (Promega) were added to a concentration of 5 mM, 0.4 mM, 0.8 U/μl, and 0.4 U/μl, respectively, in a final volume of 50 μl. The samples were incubated at 42°C for 30 min., extracted with phenol:chloroform, chloroform, precipitated and washed with ethanol, and resuspended in 25 μl of TE. This cDNA was subsequently used as a template in a PCR reaction with an adapter-specific 5' RACE inner primer (Ambion), and a gene-specific primer (BG815 for lysY, BG1088 for lysW, and BG1128 for 16S). PCR reactions with a final volume of 25 μl contained 1x RedTaq buffer (Sigma), 15 mM $MgCl_2$, 20 μM dNTP's, 50 ng of each oligo, 1.25 U of RedTaq (Sigma), $\alpha^{32}$P-dATP, and 1 μl of cDNA template. PCR reactions consisted of 35 cycle reactions of 94°C for 30 sec., 65°C for 30 sec., 72°C for 30 sec.. Radiolabeled PCR products were analyzed on a 6% acrylamide gel buffered in 1x TBE.

## *REFERENCES*


1. Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs Nucleic Acids Res. 25:3389-402.
2. Bell, S.D., A.B. Brinkman, J. van der Oost, and S.P. Jackson 2001. The archaeal TFIIEalpha homologue facilitates transcription initiation by enhancing TATA-box recognition. EMBO Rep. 2:133-8.
3. Bell, S.D., and S.P. Jackson 2001. Mechanism and regulation of transcription in archaea. Curr Opin Microbiol. 4:208-13.
4. Bell, S.D., and S.P. Jackson 2000. Mechanism of autoregulation by an archaeal transcriptional repressor. J Biol Chem. 275:31624-9.
5. Bell, S.D., and S.P. Jackson 2000. The role of transcription factor B in transcription initiation and promoter clearance in the archaeon *Sulfolobus acidocaldarius*. J Biol Chem. 275:12934-40.
6. Bell, S.D., P.L. Kosa, P.B. Sigler, and S.P. Jackson 1999. Orientation of the transcription preinitiation complex in archaea. Pro. Natl Acad Sci. 96:13662-13667.
7. Bensing, B.A., B.J. Meyer, and G.M. Dunny 1996. Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of *Enterococcus faecalis*. Proc Natl Acad Sci. 93:7794-9.
8. Borovkov, A.Y., and M.I. Rivkin 1997. XcmI-containing vector for direct cloning of PCR products. Biotechniques. 22:812-4.
9. Brinkman, A.B., I. Dahlke, J.E. Tuininga, T. Lammers, V. Dumay, E. de Heus, J.H.G. Lebbink, M. Thomm, W.M. de Vos, and J. van der Oost 2000. An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. J Biol Chem. 275:38160-38169.
10. Brinkman, A.B., T.J.G. Ettema, W.M. de Vos, and J. van der Oost 2002. The prokaryotic Lrp family of transcriptional regulators. *Submitted for publication.*
11. Calvo, J.M., and R.G. Matthews 1994. The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiol Rev. 58:466-490.
12. Chen, S., M.H. Rosner, and J.M. Calvo 2001. Leucine-regulated self-association of leucine-responsive regulatory protein (Lrp) from *Escherichia coli*. J Mol Biol. 312:625-35.
13. Ciammaruconi, A., and P. Londei 2001. *In vitro* processing of the 16S rRNA of the thermophilic archaeon *Sulfolobus solfataricus*. J Bacteriol. 183:3866-74.
14. Davies, G.E., and G.R. Stark 1970. Use of dimethyl suberimidate, a cross-linking reagent, in studying the subunit structure of oligomeric proteins. Proc Natl Acad Sci. 66:651-656.
15. Enoru-Eta, J., D. Gigot, T.L. Thia-Toong, N. Glansdorff, and D. Charlier 2000. Purification and characterization of Sa-Lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon *Sulfolobus acidocaldarius* homologous to the bacterial global transcriptional regulator Lrp. J Bacteriol. 182:3661-72.
16. Ernsting, B.R., M.R. Atkinson, A.J. Ninfa, and R.G. Matthews 1992. Characterization of the regulon controlled by the leucine-responsive regulatory protein in *Escherichia coli*. J Bacteriol. 174:1109-18.
17. Ernsting, B.R., J.W. Denninger, R.M. Blumenthal, and R.G. Matthews 1993. Regulation of the *gltBDF* operon of *Escherichia coli*: how is a leucine-insensitive operon regulated by the leucine-responsive regulatory protein? J Bacteriol. 175:7160-9.

18. Galperin, M.Y., and E.V. Koonin 1997. A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. Protein Sci. 6:2639-43.
19. Hanzelka, B.L., T.J. Darcy, and J.N. Reeve 2001. TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIEalpha. J Bacteriol. 183:1813-8.
20. Jafri, S., Evoy, S., Cho, K., Craighead, H.G., 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.
21. Kang, W.K., T. Icho, S. Isono, M. Kitakawa, and K. Isono 1989. Characterization of the gene *rimK* responsible for the addition of glutamic acid residues to the C-terminus of ribosomal protein S6 in *Escherichia coli* K12. Mol Gen Genet. 217:281-8.
22. 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. 8:123-40.
23. Kawarabayasi, Y., Y. Hino, H. Horikawa, S. Yamazaki, Y. Haikawa, K. Jin-no, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, H. Nakazawa, M. Takamiya, S. Masuda, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, H. Kikuchi, and *et al.* 1999. Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. DNA Res. 6:83-101, 145-52.
24. Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y., Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyperthermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res. 5:55-76 and 147-155.
25. Kim, J., C. Zwieb, C. Wu, and S. Adhya 1989. Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene. 85:15-23.
26. Kobashi, N., M. Nishiyama, and M. Tanokura 1999. Aspartate kinase-independent lysine synthesis in an extremely thermophilic bacterium, *Thermus thermophilus*: lysine is synthesized via alpha-aminoadipic acid not via diaminopimelic acid. J Bacteriol. 181:1713-8.
27. Kosuge, T., and T. Hoshino 1998. Lysine is synthesized through the alpha-aminoadipate pathway in *Thermus thermophilus* FEMS Microbiol Lett. 169:361-7.
28. Langer, D., Hain, J., Thuriaux, P., Zillig, W. 1995. Transcription in archaea: similarity to that in eucarya. Proc Natl Acad Sci. 92:5768-5772.
29. Leonard, P. M., S.H.J. Smits, S.E. Sedelnikova, A.B. Brinkman, W.M. de Vos, J. van der Oost, D.W. Rice, and J.B. Rafferty 2000. Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. EMBO J. 20:990-997.
30. Levinthal, M., P. Lejeune, and A. Danchin 1994. The H-NS protein modulates the activation of the *ilvIH* operon of *Escherichia coli* K12 by Lrp, the leucine regulatory protein. Mol Gen Genet. 242:736-43.
31. Lin, R., R. D'Ari, and E.B. Newman 1992. Lambda placMu insertions in genes of the leucine regulon: extension of the regulon to genes not regulated by leucine. J Bacteriol. 174:1948-55.

32. Littlefield, O., Y. Korkhin, and P.B. Sigler 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. Proc Natl Acad Sci. 96:13668-13673.
33. Madhusudhan, K.T., K.L. Hester, V. Friend, J.R. Sokatch. 1997. Transcriptional activation of the *bkd* operon of *Pseudomonas putida* by BkdR. J Bacteriol. 179:1992-1997.
34. Madhusudhan, K.T., N. Huang, and J.R. Sokatch 1995. Characterization of BkdR-DNA binding in the expression of the *bkd* operon of *Pseudomonas putida*. J Bacteriol. 177:636-641.
35. Magill, C.P., S.P. Jackson, and S.D. Bell 2001. Identification of a conserved archaeal RNA polymerase subunit contacted by the basal transcription factor TFB. J Biol Chem. 276:46693-6.
36. Marasco, R., M. Varcamonti, F. La Cara, E. Ricca, M. De Felice, and M. Sacco 1994. *In vivo* footprinting analysis of Lrp binding to the *ilvIH* promoter region of *Escherichia coli*. J Bacteriol. 176:5197-201.
37. Mathew, E., J. Zhi, and M. Freundlich 1996. Lrp is a direct repressor of the *dad* operon in *Escherichia coli*. J Bacteriol. 178:7234-7240.
38. Miyazaki, J., N. Kobashi, T. Fujii, M. Nishiyama, and H. Yamane 2002. Characterization of a *lysK* gene as an *argE* homolog in *Thermus thermophilus*. HB27 FEBS Lett. 512:269-74.
39. Miyazaki, J., N. Kobashi, M. Nishiyama, and H. Yamane 2001. Functional and evolutionary relationship between arginine biosynthesis and prokaryotic lysine biosynthesis through alpha-aminoadipate. J Bacteriol. 183:5067-73.
40. Napoli, A., J. van der Oost, C.W. Sensen, R.L. Charlebois, M. Rossi, and M. Ciaramella 1999. An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. J Bacteriol. 181:1474-1480.
41. Newman, E.B., Lin, R. 1995. Leucine-responsive regulatory protein: a global regulator of gene expression in *E.coli*. Ann Rev Microbiol. 49:747-775.
42. Nishida, H., M. Nishiyama, N. Kobashi, T. Kosuge, T. Hoshino, and H. Yamane 1999. A prokaryotic gene cluster involved in synthesis of lysine through the amino adipate pathway: a key to the evolution of amino acid biosynthesis. Genome Res. 9:1175-83.
43. Paul, L., R.M. Blumenthal, and R.G. Matthews 2001. Activation from a distance: roles of Lrp and integration host factor in transcriptional activation of *gltBDF*. J Bacteriol. 183:3910-8.
44. Platko, J.V., D.A. Willins, and J.M. Calvo 1990. The *ilvIH* operon of *Escherichia coli* is positively regulated. J Bacteriol. 172:4563-70.
45. Qureshi, S.A., S.D. Bell, and S.P. Jackson 1997. Factor requirements for transcription in the archaeon *Sulfolobus shibatae*. EMBO J. 16:2927-2936.
46. Reiter, W.D., P. Palm, W. Voos, J. Kaniecki, B. Grampp, W. Schulz, and W. Zillig 1987. Putative promoter elements for the ribosomal RNA genes of the thermoacidophilic archaebacterium *Sulfolobus sp.* strain B12. Nucleic Acids Res. 15:5581-95.
47. Ricca, E., D.A. Aker, and J.M. Calvo 1989. A protein that binds to the regulatory region of the *Escherichia coli ilvIH* operon. J Bacteriol. 171:1658-1664.
48. Roesch, P.L., and I.C. Blomfield 1998. Leucine alters the interaction of the leucine-responsive regulatory protein (Lrp) with the *fim* switch to stimulate site-specific recombination in *Escherichia coli*. Mol Microbiol. 27:751-61.
49. Sambrook, J., Fritsch, E.F., Maniatis, T. 1989. Molecular cloning, a laboratory manual, 2nd edition. Cold Spring Harbor Laboratory Press.

50. 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. 98:7835 -40.
51. Studier, F.W., A.H. Rosenberg, J.J. DunnJ.W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Meth Enzymol. 185:60 -89.
52. Tolstrup, N., C.W. Sensen, R.A. Garrett, and I.G. Clausen 2000. Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. Extremophiles. 4:175 -9.
53. van de Casteele, M., M. Demarez, C. Legrain, N . Glansdorff, A. Piérard, A. 1990. Pathways of arginine biosynthesis in extreme thermophilic archaeo - and eubacteria. J Gen Microbiology. 136:1177 -1183.
54. Weyand, N.J., B.A. Braaten, M. van der Woude, J. Tucker, and D.A. Low 2001. The essential role of t he promoter-proximal subunit of CAP in *pap* phase variation: Lrp - and helical phase -dependent activation of *papBA* transcription by CAP from –215. Mol Microbiol. 39:1504 -22.
55. Zhi, J., E. Mathew, and M. Freundlich 1999. Lrp binds to two regions in the *dadAX* promoter region of *Escherichia coli* to repress and activate transcription directly. Mol Microbiol. 32:29 -40.
56. Zhu, W., Q. Zeng, C.M. Colangelo, M. Lewis, M.F. Summers, and R.A. Scott 1996. The N -terminal domain of TFIIB from *Pyrococcus furiosus* forms a zinc ribbon. Nat Struct Biol. 3:122 -4.

# Chapter 7

## Characterization of ChoR, an archaeal-specific transcriptional regulator involved in the control of copper homeostasis

Arie B. Brinkman
Thijs J.G. Ettema
Noor G. Kornet
John van der Oost

**The archaeal transcriptional machinery, which resembles the eukaryal core system, has recently been demonstrated to be modulated by a number of bacterial-like transcriptional regulators. In this study we have analyzed ChoR, an archaeal-specific transcriptional regulator of *Sulfolobus solfataricus* that apparently functions as a copper-sensing regulator of copper homeostasis (*cho*). It is anticipated that ChoR contains an N-terminal DNA-binding helix-turn-helix domain and a C-terminal copper-binding domain. In the absence of copper, ChoR appears to repress transcription of the adjacently located *choCA* gene cluster that potentially encodes a copper chaperone (ChoC) and a P-type copper-exporting ATPase (ChoA). At growth-inhibiting copper concentrations transcription of the *choCA* cluster is induced, most likely as a result of derepression by ChoR. It is expected that the coordinated action of ChoC and ChoA gives rise to the efflux of copper, and as such to the reduction of the intracellular copper concentration to a non-toxic level. Cysteine-rich domains with similarity to that of ChoR can also be found in ChoC, some ChoA homologues, and a mammalian class of proteins that is related to myeloproliferative disease and mental retardation. Its anticipated role in copper binding is discussed.**

## INTRODUCTION

The basal transcription machinery of archaea appears to be a simplified version of the eukaryal RNA polymerase (RNAP) II system (6). It involves a eukaryal-like multi-subunit RNAP and homologues of the eukaryal general transcription factors TBP, TFIIB, and TFIIEα. Homologues of TFIIA, TFIIF, and TFIIH are not encoded by archaeal genomes, and archaeal TBP is not associated with any TBP-associated factors (TAFs) like in eukaryal TFIID (42). Besides, *in vitro* transcription from several archaeal promoters only requires TBP, TFB, and RNAP (16, 27, 28, 42). Archaeal and eukaryal RNAP II transcription initiation requires the binding of TBP (TFIID in eukarya) to a promoter element called the TATA-box, which is centered at about 25 bp upstream of the transcription start site. The resulting complex is stabilized by the binding of TFB (TFIIB in eukarya), which interacts with TBP and the TFB-responsive element (BRE), a promoter element directly upstream from the TATA-box (34). The latter interaction is mediated through a C-terminal helix-turn-helix (HTH) motif of TFB and determines transcriptional polarity (9). RNAP is subsequently recruited, which in archaea involves an interaction between the N-terminal Zn-ribbon domain of TFB and the RpoK subunit of RNAP (8, 35). Whereas the archaeal preinitiation complex (PIC) is completed at this point, the eukaryal PIC requires the additional general transcription factors TFIIE, TFIIF, and TFIIH (44).

Despite the similarities between the basal transcriptional machineries of archaea and eukarya, their regulatory mechanisms appear to be dissimilar. Although a few homologues of eukaryal-like regulators are present in archaeal genomes, many putative regulators belonging to bacterial regulatory families are encoded by archaeal genomes (see Chapter 1). For instance, most archaeal genomes encode several paralogues of the bacterial DtxR, ArsR, FUR, and Lrp family. Paralogues of Lrp are especially abundant, several of which have recently been studied (7, 10, 11, 15, 18, 36). It seems that their function is comparable in bacteria and archaea: bacterial as well as archaeal Lrp-like proteins appear to act as repressors or activators, their target genes have been shown to be involved in amino acid metabolism, and all identified ligands for Lrp-like proteins are amino acids (10, 12).

In addition to bacterial-like regulators, archaea appear to contain archaeal-specific regulators. To identify such proteins, profile search methods have been used which detect DNA-binding domains in proteins encoded by archaeal genomes (2). These analyses showed that HTH DNA-binding motifs are especially abundant, their number and diversity being comparable to that in bacteria. Besides, many Zn-ribbons are present, and to a lesser extent MetJ/Arc DNA-binding motifs (see also Chapter 1). An example of an archaeal-

specific HTH DNA-binding protein is Tfx from *Methanobacterium thermoautotrophicum*. It was shown to bind at a position 167 bp downstream the ATG start codon of its predicted target gene *fmdE*. *fmdE* is the first gene of the *fmdECB* operon that encodes molybdenum formylmethanofuran dehydrogenase. Although Tfx was proposed to be an activator of the *fmdECB* operon, direct evidence for this is still lacking, and a mechanism for activation remains to be clarified. Another example of an archaeal-specific regulator is GvpE, which activates transcription of gas-vesicle synthesis genes in halophilic archaea (23). The GvpE protein has been analyzed using molecular modeling, and was proposed to contain a leucine-zipper domain reminiscent of eukaryal bZIP transcription factors like yeast GCN4 (32, 33).

In this study we identified and characterized ChoR from *S.solfataricus*, which is an archaeal-specific DNA-binding protein. ChoR is predicted to contain an N-terminal HTH domain and a putative C-terminal copper-binding domain. The latter domain shows homology only to putative metal-binding domains of mammalian MYM-proteins (involved in myeloproliferative disease and mental retardation), to putative copper chaperones (ChoC) of archaea, and to the N-terminal domains of some archaeal and bacterial cation-transporting ATPases like ChoA. Most *choC* and *choA* genes from archaea are clustered with *cho*R. We show that transcription of *S.solfataricus choCA* is induced specifically by the addition of exogenous Cu(II) or Cd(II), suggesting that ChoC and ChoA are involved in copper and cadmium homeostasis. ChoR apparently acts as a repressor for polycistronic *choCA* transcription, since it binds to the *choCA* promoter *in vitro* only in the absence of Cu(II), conditions under which *choCA* is not transcribed *in vivo*. Cu(II) inactivates ChoR, as ChoR-DNA binding is abolished in the presence of Cu(II). The role of ChoR, ChoC, and ChoA in copper homeostasis is discussed.

## *RESULTS*

### Identification of the archaeal *choR-choCA* gene cluster and analysis of ChoR, ChoC, and ChoA

To provide a suitable model system for analyzing the function of archaeal-specific regulators, we have screened the genome of the hyperthermophilic archaeon *S.solfataricus* (47) for the presence of such proteins along with their genomic context, to enable prediction of their function and regulatory target genes. Based on the presence of a

putative HTH motif we identified a putative regulator (SSO2652) that we called ChoR (for copper homeostasis regulator).

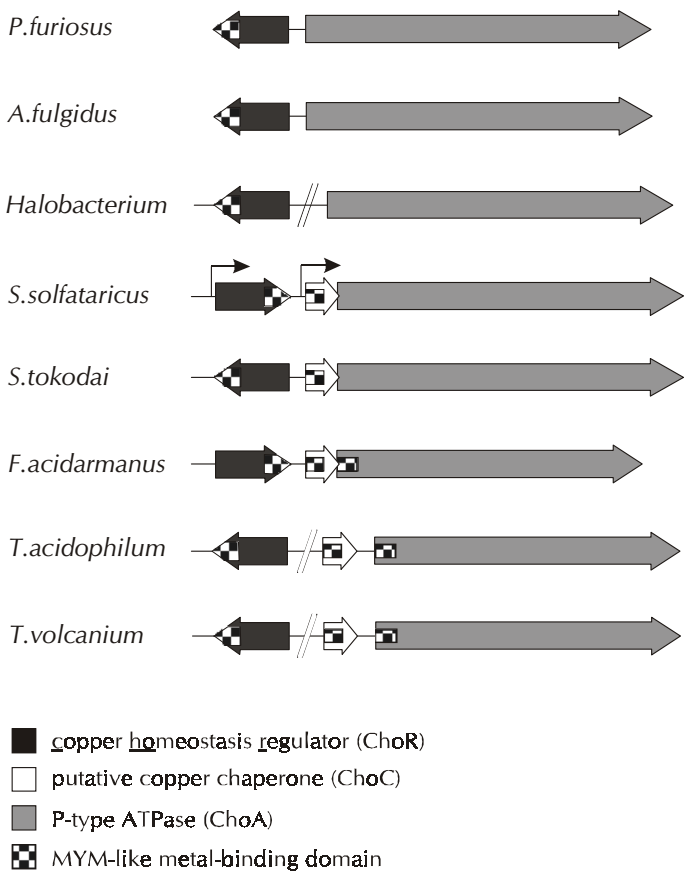



**Figure 1**. Organization of archaeal *choR-choC-choA* gene clusters. Black, *choR* genes; white, *choC* genes; and gray, *choA* P-type ATPase genes. Patterned rectangles indicate DNA sequences which deduced amino acid sequences encode the MYM-like metal-binding motif.

Genes encoding orthologues of ChoR could be found only in some archaeal genomes, but not in the genomes of bacteria or eukarya (29-31, 37, 45) (http://www.ncbi.nlm.nih.gov, http://www.jgi.doe.gov). The majority of the archaeal *cho*R genes are clustered with genes encoding a putative P-type cation-transporting ATPase (ChoA) and a small hypothetical protein that we named ChoC (see Fig. 1). P-type (phosphointermediate) ATPases form a large family of membrane proteins that that function as ATP-fuelled cation-transporting pumps (3), and ChoA of *S.solfataricus* (SSO2651) has 20 to 32% identity with several characterized copper-exporting P-type ATPases of *Enterococcus hirae* (39), *Escherichia coli* (43), *Streptococcus mutans* (49), and *Helicobacter*

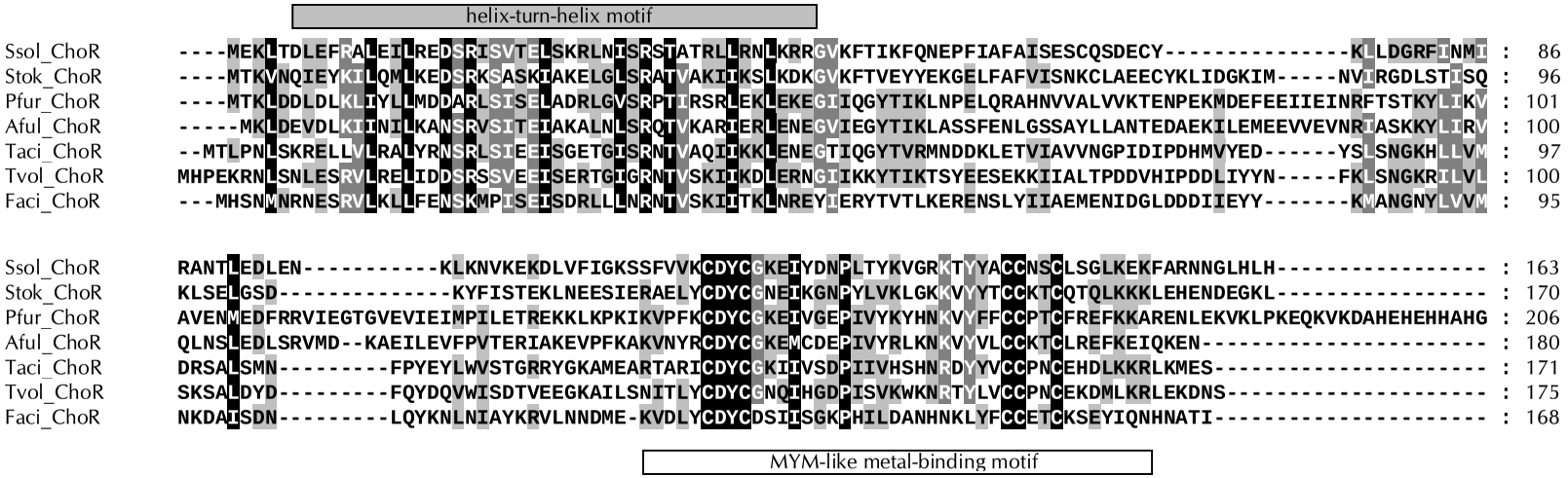



**Figure 2.** Alignment of archaeal ChoR proteins. The putative helix-turn-helix motif is indicated, as well as the MYM-like metal-binding domain. Accesion numbers: Ssol_ChoR, *Sulfolobus solfataricus* P2 Sso2652; Stok_ChoR, *Sulfolobus tokodai* ST1716; Pfur_ChoR, *Pyrococcus furiosus* AE010192; Aful_ChoR, *Archaeoglobus fulgidus* AF0474; Tac_ChoR, *Thermoplasma acidophilum* TA1218; Tvol_ChoR, *Thermoplasma volcanium* TVG0368274; Fac_ChoR, *Ferroplasma acidarmanus* contig131.revised.gene4.protein.

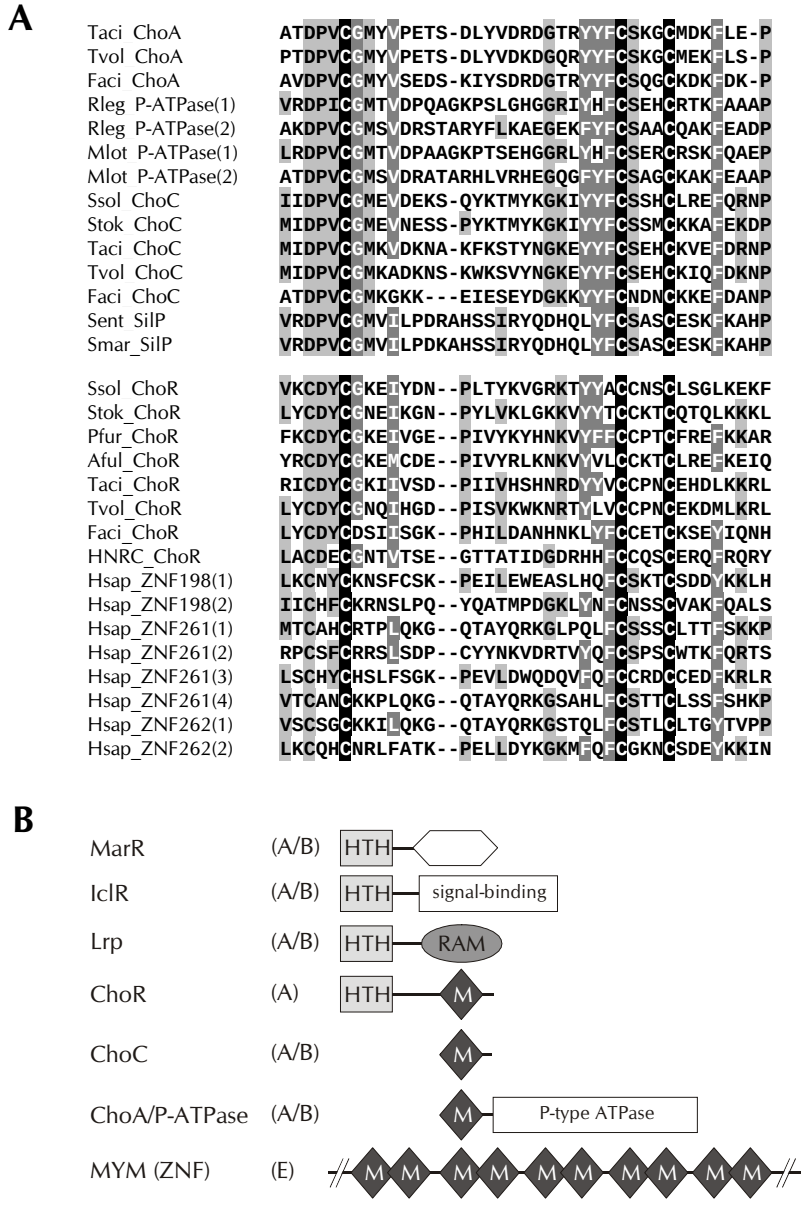



**Figure 3.** (**A**), Alignment of MYM-like metal-binding domains occurring in P-type ATPases (ChoA, SilP), ChoCs, archaeal ChoRs, and human MYM-proteins. Accession num-bers: Taci_ChoA, *T.acidophilum* TA1143; Tvol_ChoA, *T.volcanium* TVG1304145; Faci_ ChoA, *F.acidarmanus*.contig131.revised.gene6.protein; Rleg_P-ATPase, *Rhizobium leguminosarum* Q9X-5V3; Mlot_P-ATPase, *Mesorhizobium loti* Q98C24; Ssol_ ChoC, *S.solfataricus* SSO10823 ; Stok_ChoC, *S.tokodai* STS190; Taci_ChoC, *T.acidophilum* TA1144; Tvol_ChoC, *T.volcanium* TV6226; Faci_ChoC, *F.acidarmanus* contig131.revised.gene5.protein; Sent_SilP, *Salmonella enterica* AAL68937; Smar_SilP, *Serratia marcescens* AAL68932; Ssol_ChoR, *S.solfataricus* SSO2652; Stok_ChoR, *S.tokodai* ST1716; Pfur_ChoR, *P.furiosus* AE010192; Aful_ChoR, *A.fulgidus* AF0474; Taci_ChoR, *T.acidophilum* TA1218; Tvol_ChoR, *T.volcanium* TVG0368274; Faci_ChoR, *F.acidarmanus* contig131.revised.gene4.protein; HNRC_ChoR, *Halobacterium* NRC-1 VNG1179C; Hsap_ZNF198, *Homo sapiens* AL137119; Hsap_ZNF261, *H.sapiens* Q14202; Hsap_ZNF262, *H.sapiens* O43308. (**B**), Schematic structural comparison of proteins containing a HTH domain similar to that of ChoR, and proteins containing the MYM-like metal-binding domain (M). (A), archaea; (B), bacteria; (E), eukarya.

*pylori* (4), as well as with human Menkes and Wilson disease-related proteins (13). This strongly suggests that the archaeal ChoA is involved in the export of metal-ions.

SMART and BLAST analysis (1, 46), as well as HTH-prediction analysis (17) indicated that the N-terminal part of ChoR contains a HTH motif that is similar to that of several prokaryotic transcriptional regulators, such as MarR, IclR, and Lrp (Fig. 2 and 3B). Although *S.solfataricus* ChoR was annotated as an Lrp-like protein, this appears to be incorrect as its C-terminal part has no similarity with RAM domains characteristic of the C-terminus of Lrp-like proteins (12, 20) (Fig. 3B). Instead, its C-terminal part possesses a putative metal-binding motif, containing five cysteine residues located within a 30-amino acid region (Fig. 2). PSI-BLAST analysis showed that its putative metal-binding motif is conserved in all archaeal orthologues of ChoR and is furthermore characteristic of the mammalian proteins ZNF198, ZNF258, ZNF261, and ZNF262, in which multiple tandem repeats of the domain (here called Zn-finger) are present (Fig. 3) (48). The exact function of this so-called MYM family of proteins is not known, but disruption of MYM-encoding genes leads to myeloproliferative disease and mental retardation in humans. Whereas the N-terminal HTH motif and the C-terminal MYM-like metal-binding domain are well conserved between archaeal ChoRs, their central parts show a higher degree of variability (Fig. 2).

In several instances, the *choA* gene is clustered with an ORF encoding a small hypothetical protein (Fig. 1). Although the function of this protein, which we called ChoC, is not clear, PSI-BLAST analysis showed that also ChoC contains a typical cysteine motif that is also present in some archaeal and bacterial P-type ATPases, among them SilP, which is involved in silver efflux (25). Remarkably, the putative cysteine motifs of ChoCs and cation-transporting ATPases share low but significant homology with MYM-like metal-binding domains found in ChoRs and MYMs (Fig. 3). In particular, the cysteine residues that are expected to be involved in metal-binding are conserved, as well as some aromatic amino acid residues. Thus, ChoR appears to be a metal-binding regulator containing an N-terminal HTH DNA-binding domain reminiscent of bacterial regulators and a C-terminal MYM-like metal-binding motif that is also present in ChoC and some cation-transporting ATPases. The clustering of the *choR*, *choC*, and *choA* genes suggests that these genes are involved in the translocation of metal-ions and the regulation thereof.

## Expression of *choR*, *choC*, and *choA*

To determine the role of the genes in the *choR-choC-choA* locus of *S.solfataricus*, their transcription was analyzed under various conditions. First, the minimal inhibitory

| | MIC ($\mu$M) | |
|---|---|---|
| | Grogan *et al*. (24) | This study |
| EDTA | ND | < 1000 |
| $CuSO_4$ | 5000 | 5000 |
| $Ag_2SO_4$ | 8 | 8 |
| $CdCl_2$ | 2000 | > 4000 |
| $ZnSO_4$ | 50000 | > 75000 |
| $NiSO_4$ | 600 | 600 |

**Table I.** Minimal inhibitory concentrations (MICs) of various metal salts for *S.solfataricus* growth.

concentration (MIC) of different metals was determined for *S.solfataricus* growth on defined medium (Table I). The obtained MIC values were similar to the values found by Grogan *et al.* (24), except for Cd(II) and Zn(II) (Table I). The different metals were added to exponentially growing *S.solfataricus* cells on defined medium, at a concentration equivalent to the determined MIC. Total RNA was isolated and analyzed using primer extension with antisense primers for *cho*R and *choA*. As shown in Figure 4, transcription of *cho*R is constitutive under the tested conditions, whereas transcription of *choA* is induced specifically in the presence of Cu(II) and Cd(II). Under non-induced conditions *choA* appears to be transcribed at a basal level, which decreases when EDTA is added. Although addition of EDTA has no apparent effect on *cho*R transcription, it should be noted that addition of 1 mM EDTA immediately abolished *S.solfataricus* growth. Induction of *choA* transcription by Cu(II) was further studied using a time course experiment, in which *S.solfataricus* total RNA was isolated 30, 60, or 180 min. after addition of Cu(II). Transcription of *choA* was induced after 30 min. and increased during the 180 min. after induction to a level that is more than eight-fold higher than the basal transcription level (Fig. 4), as determined by quantitation of the gel. To determine whether basal *choA* transcription is due to the small amount (0.45 μM) of Cu(II) already present in the defined medium without additionally added Cu(II), we also determined *choA* transcription in defined medium in which Cu(II) is omitted. Although no significant decrease in basal transcription was observed (Fig. 4), we cannot rule out the possibility that trace amounts of Cu(II) were already present in the de-mineralized water that was used to prepare the defined medium. We conclude that the genes in the *choR-choC-choA* locus are most likely involved in the efflux of copper and cadmium, since these metals specifically induce transcription of the *choC* and *choA* genes. As with many transcriptional regulators, the *cho*R gene appears to be transcribed constitutively under the tested conditions.

To determine the transcription start sites of *cho*R and *choA*, we separated the obtained primer extension products on a larger gel along with sequence reactions that were performed with the same *cho*R and *choA* antisense primers. As shown in Fig. 4, *cho*R

transcription is initiated one basepair upstream of the start codon of the *choR* gene, whereas transcription of *choC* and *choA* is initiated two basepairs upstream of the *choC* the start codon. Since no additional bands of smaller size then shown in Figure 4 (arrow) were obtained using the *choA* antisense primer, we conclude that the *choC* and *choA* genes are transcribed as a polycistronic mRNA, initiated from a promoter upstream *choC*. Sequence elements matching *S.solfataricus* consensus TATA-boxes and BREs (9) are present upstream of the two mapped transcriptional starts (Fig. 4).

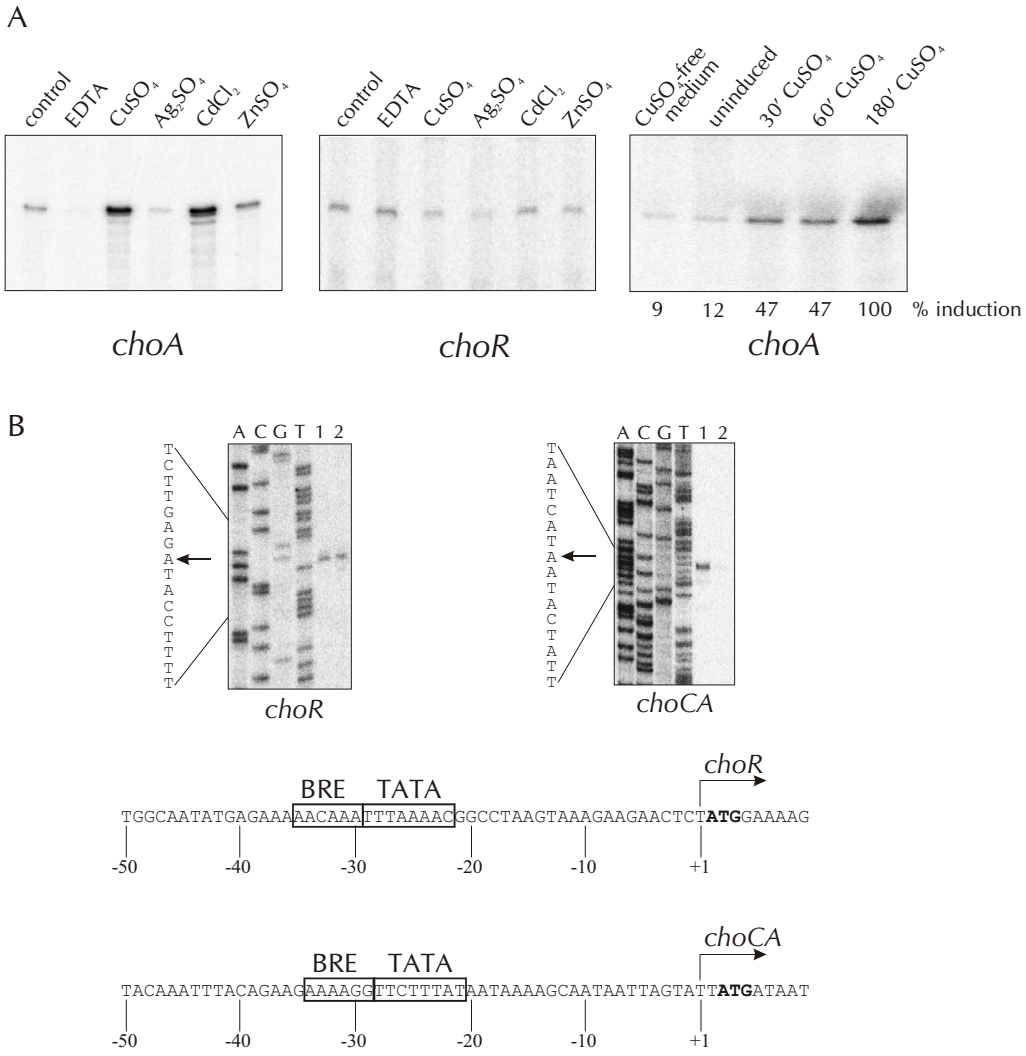



**Figure 4.** Analysis of *choR* and *choA* transcription using primer extension analysis. (**A**), Minimal inhibiting concentrations of EDTA and various metals (see Table I) were added to exponentially growing *S.solfataricus* cells. EDTA, 1 mM; $CuSO_4$, 5 mM; $Ag_2SO_4$, 8 $\mu$M; $CdCl_2$, 4 mM; $ZnSO_4$, 75 mM. Three hours after addition of EDTA or metals total RNA was isolated and analyzed using primer extension with *choR* and *choA* antisense primers. *choA* transcript levels of a $CuSO_4$ time-course experiment (right) were quantified using a phosphorimager. (**B**), Mapping of the *choR* and *choCA* transcription start sites. 1, RNA from $CuSO_4$-induced cells; 2, RNA from non-induced cells. The transcriptional start sites are indicated with an arrow (+1), and presumptive BREs and TATA-boxes are rectangled.

## ChoR binds to the *choCA* promoter in the absence of metal-ions

The *choR* gene was cloned into pET24d, and functionally expressed in *E.coli* BL32(λDE3)-pRIL. Purification was most efficient when ChoC was expressed as a C-terminally His-tagged protein, and Ni-affinity chromatography was performed under non-denaturing conditions. The purified ChoR was used in electrophoretic mobility shift assays (EMSAs), to determine whether it binds to the mapped *choR* and *choCA* promoters. ChoR did not bind to a DNA fragment called $P_{choR}$ (Fig. 5A), containing the *choR* promoter and its -150 to +50 flanking sequence (not shown). In contrast, binding of ChoR was observed to $P_{choC}$, containing the *choC* promoter and its -120 to +130 flanking sequence (Fig. 5A and B). The EMSA revealed that ChoR binding to $P_{choC}$ was relatively weak, and that three ChoR-DNA complexes of distinct electrophoretic mobility were formed (Fig. 5B, I, II, and III). Because ChoR is expected to be a metal-responsive regulator, metal-ions that may bind with high affinity to ChoR and potentially inhibit DNA-binding are possibly not removed during protein purification under native conditions. Therefore, we tested the effect of EDTA in the binding reaction. As shown in Figure 5B, scavenging such metal-ions with EDTA drastically increased ChoR-DNA binding affinity, and quantitation of the image revealed that the increase in affinity was at least 100-fold (not shown).

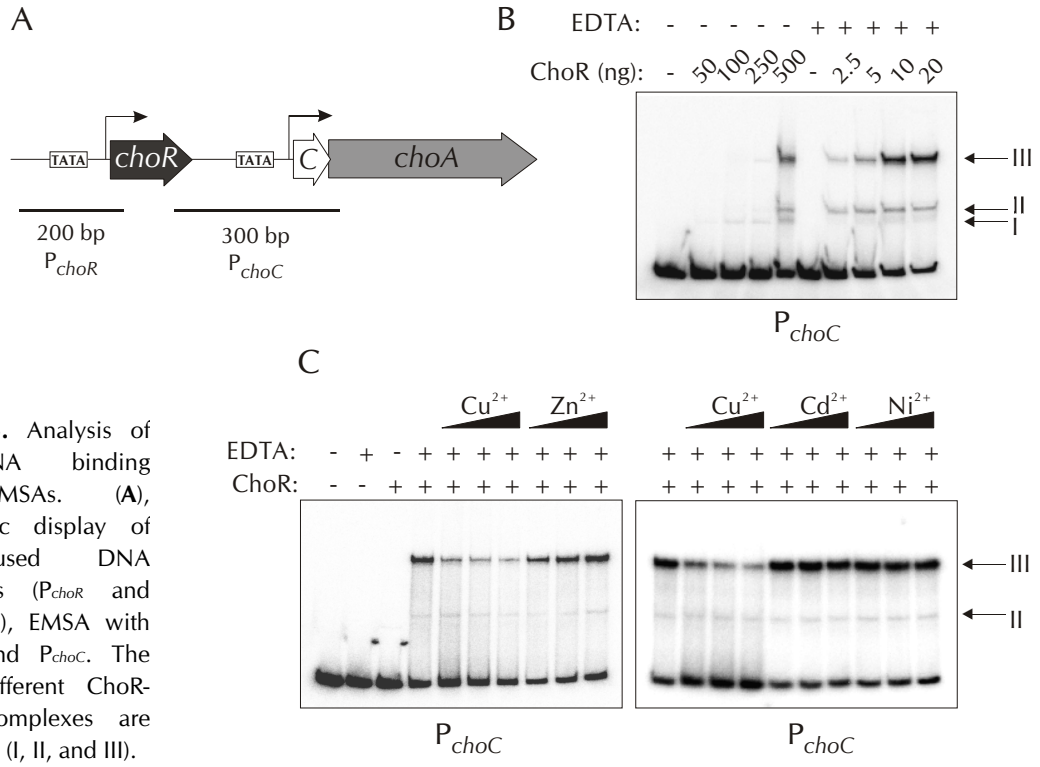



**Figure 5.** Analysis of ChoR-DNA binding using EMSAs. (**A**), Schematic display of the used DNA fragments ($P_{choR}$ and $P_{choC}$). (**B**), EMSA with ChoR and $P_{choC}$. The three different ChoR-DNA complexes are indicated (I, II, and III). A final concentration of 50 mM of EDTA was present in the binding reactions as indicated. (**C**), Effect of various metal-ions on ChoR-DNA binding in the presence of 50 mM EDTA. Metal salts were added to a final concentration of 20, 25, and 30 mM, respectively.

### Cu(II) decreases the affinity of ChoR for *choCA* promoter DNA

To identify the metal-ions that inhibit ChoR DNA-binding, we performed EMSAs with $P_{choC}$ and ChoR in the presence of EDTA, and tested the effect of several metal-ions in the binding reaction (Fig. 5C). Of the four metal-ions tested (Cu(II), Zn(II), Cd(II), Ni(II)), only Cu(II) inhibited ChoR-DNA binding. Together with the observed induction of *choA* transcription by exogenous Cu(II), this suggests that ChoR is a repressor of *choCA* transcription. No effect was observed when Cd(II) was added to the binding reaction.

## *DISCUSSION*

While copper is an essential element for living organisms, cytoplasmic copper overexposure is highly toxic due to the generation of hydroxyl radicals which cause severe damage to DNA, proteins, and lipids. Protection against copper overexposure in bacteria and eukarya involves P-type cation-transporting ATPases that export excessive copper-ions (40). In this study we have identified an archaeal gene cluster which is obviously involved in copper and possibly cadmium homeostasis through the use of a similar type ATPase. Transcriptional regulation of this archaeal copper export system involves a novel archaeal-specific regulator, containing a N-terminal HTH DNA-binding domain, and a C-terminal cysteine-rich metal-binding domain. While transcription of *choC* and *choA* is induced upon addition of Cu(II) or Cd(II) to the medium, Cu(II) specifically affects DNA-binding of ChoR: the presumptive apo-ChoR binds the *choCA* promoter with at least 100-fold higher affinity than the putative Cu(II)-ChoR. Hence, it is anticipated that under non-inducing conditions binding of ChoR causes repression of the *choCA* promoter, which is relieved under inducing conditions.

Because transcriptional analysis showed *choCA* induction by both Cu(II) and Cd(II), ChoR-DNA binding was expected to be responsive to both metals. However, we only detected an obvious effect of Cu(II). In contrast to other metalloregulatory proteins that have various metal-ion ligands (5), ChoR thus seems to be more specific to only one metal-ion. It is not clear how *choCA* transcription is induced by Cd(II), but it may involve an additional, yet unidentified, regulator. Since our preparation of ChoR was apparently not totally free of metal-ions, EMSAs required the presence of EDTA in the binding reaction. Our observation that ChoR is not responsive to Cd(II) might therefore be an experimental artifact. Obviously, metal-ions have become complexed with ChoR either during expression in *E.coli* or during subsequent purification of the protein, where traces of metal

salts might have been present in the used buffers. Like other metalloregulatory proteins, ChoR thus binds its metal-ligand with high affinity, and denaturation-renaturation of such protein preparations in the presence of EDTA is sometimes necessary to obtain the apo-form of these proteins (S.D. Bell, personal communication). Unfortunately, such a procedure was unsuccessful for ChoR. In addition, it should be noted that efficient purification of ChoR required the presence of a C-terminal $His_6$-tag, which may interfere with or alter the metal-response compared to the wild-type ChoR.

Because ChoR-DNA binding experiments required the presence of EDTA, DNA footprinting experiments could not be employed, as the various DNA cleavage agents used for footprinting are metal-complexes or require metal-ions as co-factors. To identify a possible ChoR binding site, we aligned the *S.solfataricus choC-cta* promoter with the putative analogous promoters of other archaea, however, no homologous sequence elements were present. Even the promoters of the most closely related organisms *S.solfataricus* and *S.tokodai* share no obvious homology in their core promoter elements (not shown).

Although copper is a highly toxic element to cellular components, many cytoplasmic enzymes do require the incorporation of copper for their catalytic function. Hence, intracellular trafficking of copper and prevention against cytoplasmic copper exposure in bacteria and eukarya is performed by copper chaperones (26). Although a small number of copper chaperones have been identified, it seems that each cellular copper-dependent protein is served by a specific copper chaperone. For example, copper-export by the P-type ATPases CopA of *E.hirae* and CCC2 of *S.cerevisiae* requires the specific copper chaperones CopZ and ATX1, respectively (14, 38, 41). These copper chaperones are small proteins with an N-terminal MxCxxC-motif involved in the binding of Cu(I), as determined by structural studies (19, 22). The MxCxxC-motif is not only present in the N-terminal part of these copper chaperones, but also in the N-terminal region of their respective P-type ATPases and other P-type ATPases, sometimes in multiple copies (Fig. 6), although their role here is not completely clear, since mutation of the cysteine residues does not affect the resistance to and the transport of copper (21). For several reasons, it is possible that archaeal ChoC functions as a copper chaperone for ChoA. First, because its gene is clustered and co-transcribed with that of *choA*, and induced by the presence of exogenous copper. Second, because ChoC contains the MYM-like metal-binding motif that is also present N-terminally in some archaeal and bacterial P-type ATPases. This motif may be a functional analogue of the MxCxxC motif. Interestingly, some archaeal CTAs contain the MYM-like motif, whereas others contain the MxCxxC-motif. In other words, the presence of either the MYM-like motif or the MxCxxC-motif in CTAs appears to be mutually

exclusive. Like the human Menkes/Wilson proteins and their MxCxxC-motifs, human MYM-proteins contain multiple copies of the MYM-like metal-binding domain (48), and in analogy, they may play a role in metal-homeostasis. Further analyses are required to elucidate the role and function of these MYM-like metal-binding motifs.

In conclusion, we have identified an archaeal gene cluster involved in the homeostasis of copper (and cadmium). Its regulation involves a novel archaeal-specific DNA-binding protein containing a bacterial-like HTH domain and a MYM-like metal-binding domain. The latter domain appears to play a central role in the homeostasis of copper and cadmium in *S.solfataricus*, since it is present in all three involved proteins.

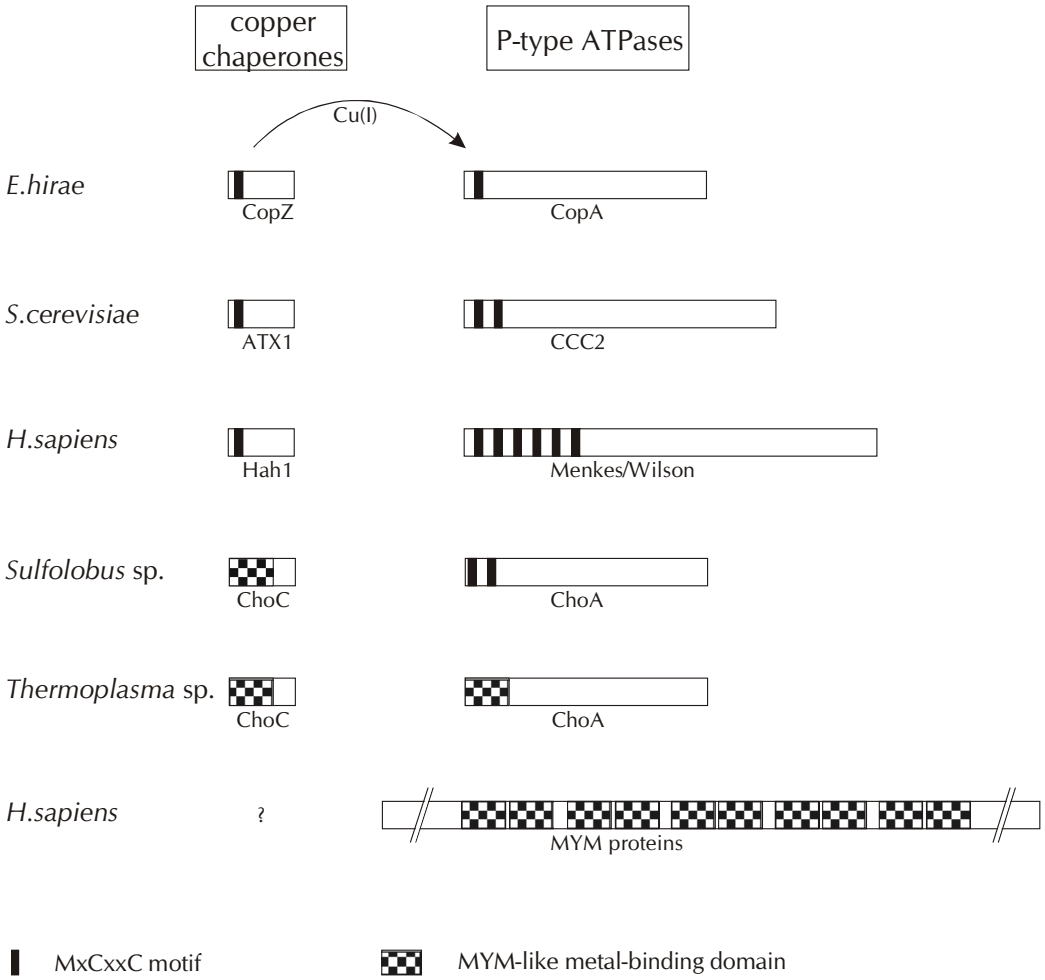



**Figure 6.** Schematic structural comparison of copper exporting P-type ATPases and their specific copper chaperones in *Enterococcus hirae*, *Saccheromyces cerevisiae*, *Homo sapiens*, and the putative copper chaperones of archaea. Filled bars indicate MxCxxC motifs, and patterned rectangles indicate MYM-like metal-binding domains. Accession numbers: *E.hirae* CopZ, Q47840; *E.hirae* CopA, P32113; *S.cerevisiae* ATX1, P38636; *S.cerevisiae* CCC2, P38995; *H.sapiens* Hah1, O00244; *H.sapiens* Menkes (AT7A), Q04656; *H.sapiens* Wilson (AT7B), P35670. See Fig. 3 for other accession numbers.

## ACKNOWLEDGEMENTS

We thank S.D. Bell for helpful experimental advice. This research was supported by the Council for Chemical Sciences (CW) of the Netherlands Organization for Scientific research (NWO) grant 700-35-101.

## EXPERIMENTAL PROCEDURES

### Growth of *S.solfataricus*

Growth of *S.solfataricus* in defined medium was performed as described (10). For determination of the MIC values of the different metal salts, various concentrations of these salts were added to exponentially growing *S.solfataricus* cells ($OD_{600}$ of 0.3-0.4), and growth was monitored by measuring the $OD_{600}$.

### RNA isolation and primer extension analysis

RNA isolation and primer extension analysis were performed as described in (10). The antisense primers used for primer extension were BG1131 (5'-GTGCTCCTACTG-ATATTAAGCC-3') for *choR*, and BG1130 (5'-CATGTTGCACAATGCATCCC-3') for *choA*.

### EMSAs

Electrophoretic mobility shift assays (EMSAs) were performed as described (10), with minor adjustments. The binding buffer consisted of 50 mM Tris pH 8.0, 5 mM DTT, 10% glycerol, and 5 ng/µl poly(dI.dC).poly(dI.dC). EDTA, if present, was added to a final concentration of 50 mM. The DNA fragments used were generated using PCR with the primers BG1081 (5'-ACTAGTTGGATGGATATTAGGAATAGC-3') and BG1082 (5'-TCTCTTAAAATCTCCAGCGCTC-3') for $P_{choR}$, and primers BG1079 (5'-TGCACGCAA-CAATGGCTTGC-3') and BG1080 (5'-CTTTCTTCTCAGTGGCCATGTG-3') for $P_{choC}$.

### Production and purification of ChoR

The *choR* gene was PCR-amplified from *S.solfataricus* genomic DNA using the primers BG864 (5'-CGCGCCATGGAAAAGTTGACAGATTTAGAGTTTAG-3') and BG1120 (5'- CGCGCCTCGAGATGTAAGTGCAAGCCATTGTTGCG-3'), which contain *Nco*I and *Xho*I restriction sites, respectively (underlined), and generated a PCR fragment that was

cloned into *Nco*I-*Xho*I digested pET24d (Novagen) resulting in pWUR58. This construct was subsequently transformed to *E.coli* BL21(λDE3)-pRIL (Novagen) to produce C-terminally 6xHis-tagged ChoR protein. *E.coli* cells harboring pWUR58 were grown at 37°C in one liter of LB medium to an $OD_{600}$ of 0.5, and ChoR expression was induced by the addition of IPTG to a final concentration of 0.1 mM. After an additional four-hour incubation at 37°C the cells were harvested by centrifugation for 10 min. at 6000 RPM. A bacterial cell pellet derived from 250 ml of culture was resuspended in 20 ml of 50 mM Tris pH 8.0, 10 mM β-mercaptoethanol. Fractions of 1 ml were sonicated three times for 15 seconds on ice and centrifuged for 10 min. at 13.000 RPM. To the pooled supernatant, NaCl and imidazole was added to a final concentration of 300 mM of 10 mM, respectively, and the supernatant was subsequently loaded on a chelating sepharose column (Amersham-Pharmacia) loaded with $NiSO_4$. After washing the column with three column volumes of the same buffer, ChoR was eluted with a pulse of one column volume of the same buffer containing 1 M of imidazole, and 1-ml fractions containing ChoR were collected.

## REFERENCES


1. Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-402.
2. Aravind, L., Koonin, E.V. 1999. DNA-binding proteins and evolution of transcription regulation in the archaea. Nucleic Acids Res 27:4658-4670.
3. Axelsen, K.B., and M.G. Palmgren 1998. Evolution of substrate specificities in the P-type ATPase superfamily. J Mol Evol 46:84-101.
4. Bayle, D., S. Wangler, T. Weitzenegger, W. Steinhilber, J. Volz, M. Przybylski, K.P. Schafer, G. Sachs, and K. Melchers 1998. Properties of the P-type ATPases encoded by the *copAP* operons of *Helicobacter pylori* and *Helicobacter felis*. J Bacteriol 180:317-29.
5. Bell, S.D., S.S. Cairns, R.L. Robson, and S.P. Jackson 1999. Transcriptional regulation of an archaeal operon *in vivo* and *in vitro*. Mol Cell 4:971-982.
6. Bell, S.D., and S.P. Jackson 2001. Mechanism and regulation of transcription in archaea. Curr Opin Microbiol 4:208-13.
7. Bell, S.D., and S.P. Jackson 2000. Mechanism of autoregulation by an archaeal transcrip tional repressor. J Biol Chem 275:31624-9.
8. Bell, S.D., and S.P. Jackson 2000. The role of transcription factor B in transcription initiation and promoter clearance in the archaeon *Sulfolobus acidocaldarius*. J Biol Chem 275:12934-40.
9. Bell, S.D., P.L. Kosa, P.B. Sigler, and S.P. Jackson 1999. Orientation of the transcription preinitiation complex in archaea. Proc Natl Acad Sci U S A 96:13662-13667.
10. Brinkman, A.B., Bell, S.D., Lebbink, R.J., de Vos, W.M., van der Oost, J. 2002. The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. J Biol Chem *In press*.
11. Brinkman, A.B., I. Dahlke, J.E. Tuininga, T. Lammers, V. Dumay, E. de Heus, J.H.G. Lebbink, M. Thomm, W.M. de Vos, and J. van der Oost 2000. An Lrp-like transcriptional regulator from the archaeon Pyrococcus furiosus is negatively autoregulated. J Biol Chem 275:38160-38169.
12. Brinkman, A.B., T.J.G. Ettema, W.M. de Vos, and J. van der Oost 2002. The prokaryotic Lrp family of transcriptional regulators. *Manuscript in preparation*.
13. Bull, P.C., G.R. Thomas, J.M. Rommens, J.R. Forbes, and D.W. Cox 1993. The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. Nat Genet 5:327-37.
14. Cobine, P., W.A. Wickramasinghe, M.D. Harrison, T. Weber, M. Solioz, and C.T. Dameron 1999. The *Enterococcus hirae* copper chaperone CopZ delivers copper(I) to the CopY repressor. FEBS Lett 445:27-30.
15. Dahlke, I., and M. Thomm 2002. A *Pyrococcus* homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment. Nucleic Acids Res 30:701-10.
16. Darcy, T.J., W. Hausner, D.E. Awery, A.M. Edwards, M. Thomm, and J.N. Reeve 1999. *Methanobacterium thermoautotrophicum* RNA polymerase and transcription *in vitro*. J Bacteriol 181:4424-9.
17. Dodd, I.B., and J.B. Egan 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res 18:5019-26.
18. Enoru-Eta, J., D. Gigot, T.L. Thia-Toong, N. Glansdorff, and D. Charlier 2000. Purification and characterization of Sa-Lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon

*Sulfolobus acidocaldarius* homologous to the bacterial global transcriptional regulator Lrp. J Bacteriol 182:3661-72.

19. Eriksson, P.O., and L. Sahlman 1993. 1H NMR studies of the mercuric ion binding protein MerP: sequential assignment, secondary structure and global fold of oxidized MerP. J Biomol NMR 3:613 -26.
20. Ettema, T.J.G., A.B. Brinkman, P.M. Leonard, T.H. Tani, J.B. Rafferty, and J. van der Oost 2002. A novel ligand-binding domain (RAM) involved in allosteric regulation of amino acid metabolism in prokaryotes. J Biol Chem, *in press*.
21. Fan, B., G. Grass, C. Rensing, and B.P. Rosen 2001. *Escherichia coli* CopA N-terminal Cys(X)(2)Cys motifs are not required for copper resistance or transport. Biochem Biophys Res Commun 286:414 -8.
22. Gitschier, J., B. Moffat, D. Reilly, W.I. Wood, and W.J. Fairbrother 1998. Solution structure of the fourth metal-binding domain from the Menkes copper-transporting ATPase. Nat Struct Biol 5:47 -54.
23. Gregor, D., and F. Pfeifer 2001. Use of a halobacterial *bgaH* reporter gene to analyse the regulation of gene expression in halophilic archaea. Microbiology 147:1745 -54.
24. Grogan, D.W. 1989. Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. J Bacteriol 171:6710-9.
25. Gupta, A., K. Matsui, J.F. Lo, and S. Silver 1999. Molecular basis for resistance to silver cations in *Salmonella*. Nat Med 5:183-8.
26. Harrison, M.D., C.E. Jones, and C.T. Dameron 1999. Copper chaperones: function, structure and copper-binding properties. J Biol Inorg Chem 4:145 -53.
27. Hausner, W., J. Wettach, C. Hethke, M. Thomm 1996. Two transcription fa ctors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. J Biol Chem 271:30144 -30148.
28. Hethke, C., A.C. Geerling, W. Hausner, W.M. de Vos, and M. T homm 1996. A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. Nucleic Acids Res 24:2369-2376.
29. 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 genom e sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. DNA Res 8:123-40.
30. Kawashima, T., N. Amano, H. Koike, S. Makino, S. Higuchi, Y. Kawashima -Ohya, K. Watanabe, M. Yamazaki, K. Kanehori, T. Kawamoto, T. Nunoshiba, Y. Ya mamoto, H. Aramaki, K. Makino, and M. Suzuki 2000. Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. Proc Natl Acad Sci U S A 97:14257-62.
31. Klenk, H.P., R.A. Clayton, J.F. Tomb, O. White, K.E. Nelson, K.A . Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, D.L. Richardson, A.R. Kerlavage, D.E. Graham, N.C. Kyrpides, R.D. Fleischmann, J. Quackenbush, N.H. Lee, G.G. Sutton, S. Gill, E.F. Kirkness, B.A. Dougherty, K. McKenney, M.D. Adams, B. Loftus, J.C. Venter, and *et al.* 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature 390:364-370.
32. Konig, P., and T.J. Richmond 1993. The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. J Mol Biol 233:139 -54.

33. Krüger, K., T. Hermann, V. Armbruster, F. Pfeifer, F. 1998. The transcriptional activator GvpE for the halobacterial gas vesicle genes resembles a basic region leucine -zipper regulatory protein. J Mol Biol 279:761-771.
34. Littlefield, O., Y. Korkhin, and P.B. Sigler 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. Proc Natl Acad Sci U S A 96:13668 -13673.
35. Magill, C.P., S.P. Jackson, and S.D. Be ll 2001. Identification of a conserved archaeal RNA polymerase subunit contacted by the basal transcription factor TFB. J Biol Chem 276:46693 -6.
36. Napoli, A., J. van der Oost , C.W. Sensen, R.L. Charlebois, M. Rossi, and M. Ciaramella 1999. An Lrp -like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. J Bacteriol 181:1474-1480.
37. Ng, W.V., S.P. Kennedy, G.G. Mahairas, B. Berquist, M. Pan, H.D. Shukla, S.R. Lasky, N.S. Baliga, V. Thorsson, J. Sbrogna, S. Swartzell, D. Weir, J. Hall, T.A. Dahl, R. Welti, Y.A. Goo, B. Leithauser, K. Keller, R. Cruz, M.J. Danson, D.W. Hough, D.G. Maddocks, P.E. Jablonski, M.P. Krebs, C.M. Angevine, H. Dale, T.A. Isenbarger, R.F. Peck, M. Pohlschroder, J.L. Spudich, K.W. Jung, M. Alam, T. Freitas, S. Hou, C.J. Daniels, P.P. Dennis, A.D. Omer, H. Ebhardt, T.M. Lowe, P. Liang, M. Riley, L. Hood, and S. DasSarma 2000. Genome sequence of *Halobacterium* species NRC-1. Proc Natl Acad Sci U S A 97:12176-81.
38. Odermatt, A., and M. Soli oz 1995. Two trans-acting metalloregulatory proteins controlling expression of the copper-ATPases of *Enterococcus hirae*. J Biol Chem 270:4349-54.
39. Odermatt, A., H. Suter, R. Krapf, and M. Solioz 1993. Primary structure of two P -type ATPases involved in copper homeostasis in *Enterococcus hirae*. J Biol Chem 268:12775-9.
40. Pena, M.M., J. Lee, and D.J. Thiele 1999. A delicate balance: homeostatic control of copper uptake and distribution. J Nutr 129:1251-60.
41. Pufahl, R.A., C.P. Singer, K.L. Peariso, S.J . Lin, P.J. Schmidt, C.J. Fahrni, V.C. Culotta, J.E. Penner -Hahn, and T.V. O'Halloran 1997. Metal ion chaperone function of the soluble Cu(I) receptor Atx1. Science 278:853-6.
42. Qureshi, S.A., S.D. Bell, and S.P. Jackson 1997. Factor requirements for tra nscription in the archaeon *Sulfolobus shibatae*. EMBO J 16:2927-2936.
43. Rensing, C., B. Fan, R. Sharma, B. Mitra, and B.P. Rosen 2000. CopA: An *Escherichia coli* Cu(I)-translocating P-type ATPase. Proc Natl Acad Sci U S A 97:652 -6.
44. Roeder, R.G. 1996. The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci 21:327-335.
45. Ruepp, A., W. Graml, M.L. Santos -Martinez, K.K. Koretke, C. Volker, H.W. Mewes, D. Frishman, S. Stocker, A.N. Lupas, and W. Baumeister 2000. The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. Nature 407:508-13.
46. Schultz, J., F. Milpetz, P. Bork, and C.P. Ponting 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95:5857-64.
47. 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 U S A 98:7835-40.

48. Smedley, D., R. Hamoudi, Y.J. Lu, C. Cooper, and J. Shipley 1999. Cloning and mapping of members of the MYM family. Genomics 60:244-7.
49. Vats, N., and S.F. Lee 2001. Characterization of a copper-transport operon, *copYAZ*, from *Streptococcus mutans*. Microbiology 147:653-62.

# *Chapter 8*

Summary and concluding remarks

Archaea comprise the third domain of life and, although they resemble bacteria with respect to their prokaryotic cellular organization, several essential molecular processes like transcription resemble that of eukarya. The basal archaeal transcription machinery has been studied extensively in the past two decades. However, little is known about archaeal mechanisms of transcriptional regulation. The research described in this thesis is aimed to address the characterization of selected regulatory proteins involved in archaeal transcription, as well as the mechanisms by which these regulators influence transcription.

In **Chapter 1** the characteristics of the archaeal transcription machinery are described and compared to that of bacteria and eukarya. Moreover, various transcription-related aspects are discussed. The basal components involved in archaeal transcription initiation (TBP, TFB, RNAP) resemble those of eukarya and function in a highly similar way. In addition, three of the four eukaryal core promoter elements are also present in archaeal promoters. Except for TFIIE, archaeal genomes do not encode homologues of the numerous additional transcription factors that are required for eukaryal transcription initiation. Archaeal TFE is homologous to the N-terminal region of the eukaryal TFIIE α-subunit, and corresponds to the minimal TFIIE fragment that supports yeast growth. The characterization of *S.solfataricus* TFE is described in **Chapter 2**, and it is shown that while TFE is not absolutely required for transcription in a reconstituted *in vitro* system, it nonetheless plays a stimulatory role when TBP is limiting or at promoters containing sub-optimal TATA-box sequences, suggesting that TFE facilitates or stabilizes the TBP-TATA-box interaction.

The use of bioinformatics in the analysis of several completed archaeal genome sequences has facilitated the identification of archaeal genes encoding regulators. It has been demonstrated that although the archaeal transcription apparatus is eukaryal-like, most of the present regulators are bacterial-like. Members of the Lrp family and the metal-responsive regulatory families ArsR, DtxR, and FUR are ubiquitously present in archaeal genomes. It is anticipated that archaeal genomes do not only encode bacterial-like regulators, but also archaeal-specific regulators. Profile search analysis of archaeal genomes has indicated the presence of numerous putative DNA-binding domains. Whereas the majority of these domains are HTH domains, MetJ/Arc-domains and Zn-ribbon-domains are present as well. The most abundant class of bacterial-like regulators encountered in archaea is that of Lrp, named after the leucine-responsive regulatory protein of *E.coli*, which acts as a transcriptional regulator for many genes involved mainly in amino acid metabolism, but also in other essential processes. In **Chapter 3** the properties of bacterial and archaeal Lrp-like proteins are reviewed. In contrast to *E.coli* Lrp, which acts as a global

regulator, other bacterial Lrp-like proteins appear to be specific regulators for one gene or operon. The number of genes encoding Lrp-like proteins is comparable for bacterial and archaeal genomes, and varies from 1 to 10 per genome. The presently available data suggests that Lrp-like proteins are generally involved in the regulation of amino acid metabolism. This is in agreement with the fact that all identified ligands for Lrp-like proteins are amino acids.

The analysis of the LrpA structure (see below) is discussed in detail. Although the LrpA helix-turn-helix (HTH) domain is homologous to HTHs encountered in bacterial and eukaryal DNA-binding proteins, superimposition shows that its fold is most similar to that of the bacterial catabolite activating protein (CAP). Comparison of the LrpA HTH with the HTH of CAP in complex with DNA allows the prediction of the LrpA amino acid residues that make base-specific DNA contacts. However, which of the bases in the LrpA binding site are contacted by these residues remains unclear, and requires the co-crystallization of LrpA and its target sequence. Examination of the LrpA C-terminal domain shows that its structure resembles that of an aspartate kinase-chorismate mutase-TyrA (ACT) domain with respect to structure and function: both consist of the βαββαβ-fold, and are regulatory domains involved in allosteric modulation of the activity of enzymes and regulators involved in amino acid metabolism. Because of the general physiological role of Lrp-like proteins as regulators of amino acid metabolism, we propose to refer to this novel regulatory module as RAM domain. It is striking that although the structures and functions of ACT and RAM are similar, they differ with respect to their amino acid sequence and their anticipated effector binding sites. It is concluded that the analogy between ACT and RAM is the result of convergent evolution.

The role and function of two archaeal Lrp-like proteins was studied in detail. The first of the two is LrpA from *P.furiosus*, the gene of which had been identified previously during isolation and sequencing of the gene encoding glutamate dehydrogensase (*gdh*). The characterization of LrpA is described in **Chapter 4**. The *lrpA* gene is located downstream of *gdh*, and since GDH plays an important role in amino acid metabolism, it was anticipated that LrpA might act as a regulator for *gdh* transcription. *In vitro* transcription assays showed that, although LrpA has no apparent effect on *gdh* transcription, it acts as a negative regulator for transcription of the *lrpA* gene itself. As for most bacterial Lrp-like proteins, LrpA autoregulation is ligand-independent. In agreement with this, purified LrpA binds to the *lrpA* promoter, to a site downstream of the TATA-box and centered at the transcription start site of *lrpA*, as determined by electrophoretic gel-mobility shift assays and different DNA footprinting techniques. The LrpA binding site is at least 43 bp long,

and mutational analysis of this site showed that a GGTTC-element is specifically recognized and required for efficient binding and repression by LrpA. Although LrpA appears to exist as a dimer, tetramer, and octamer in solution, chemical cross-linking analysis showed that LrpA assembles preferably into a tetrameric protein, both in solution and bound to the DNA. The location of the LrpA binding site suggests that the binding of TATA-binding protein (TBP) and transcription factor B (TFB) is not blocked by LrpA, which suggests that the mechanism of LrpA repression is prevention of RNA polymerase (RNAP) recruitment.

Because no 3D structure of any Lrp-like protein was available, crystallization experiments of the efficiently produced and purified of *P.furiosus* LrpA were initiated. We successfully obtained LrpA crystals and were able to solve the LrpA structure at a resolution of 2.9 Å, as described in **Chapter 5**. Its structure comprises an N-terminal HTH DNA-binding domain connected with a hinge to a C-terminal domain consisting of two α-helices packed on one side of a four-stranded anti-parallel β-sheet. The two domain-structure of LrpA are is in agreement with an earlier mutational analysis of *E.coli* Lrp, which showed that DNA-binding activity resides in the N-terminal part of the protein, whereas transcriptional activation and ligand response is confined to the central and C-terminal part of the protein.

The second archaeal Lrp-like protein studied in this thesis is LysM from *S.solfataricus*. Because the physiological function of the many archaeal Lrp-like proteins is unknown, we attempted to identify Lrp-homologues for which target genes other than its own coding gene could more easily be predicted, using genomic context analysis. In **Chapter 6** the identification and characterization of LysM is described. Its gene is clustered with *lys* genes that encode a (novel) prokaryotic lysine biosynthesis pathway so far only encountered in the thermophilic bacterium *T.thermophilus* and several (mostly hyperthermophilic) archaea. The *lys* operon is transcribed from two different promoters. The first promoter ($P_{lysY}$) is located upstream of *lysYZM* and is constitutive, whereas the second ($P_{lysW}$) is located upstream of *lysWXJK*, and is highly regulated by lysine availability in the medium, with maximum activity in the absence of lysine. Mobility shift assays and DNaseI footprinting experiments showed that purified LysM binds to the $P_{lysW}$ promoter, at a binding site located directly upstream of the TFB-responsive element (BRE) of $P_{lysW}$. A conserved inverted repeat element at this position appears to be recognized by LysM. In agreement with the anticipated role of LysM in lysine biosynthesis, its DNA-binding affinity is reduced specifically by lysine. Although evidence is provided that $P_{lysW}$ functions as a promoter *in vivo*, its activity could not be measured *in vitro*, which raises the possibility that a co-activator

is required for activation of this intrinsically weak promoter. In a model for regulation of the $P_{lysW}$ promoter based on our results, we conclude that both maximal promoter activity and highest LysM-DNA binding affinity are preferred in the absence of lysine, suggesting that LysM acts as an activator of transcription. If this proves to be the case, LysM would be the first example of a bacterial-like regulator that is able to activate the eukaryal-like transcription machinery of archaea.

The characterization of LysM and LrpA demonstrates that the role of Lrp-like proteins from bacteria and archaea is very similar: they are involved in the regulation of amino acid metabolism, and can act as potential repressors and activators. Furthermore, the ligand-response is comparable: the induced changes in DNA-binding affinity are subtle rather than absolute.

It is anticipated that archaeal genomes encode a number of archaeal-specific regulators that can potentially be identified on the basis of a present DNA-binding domain. In **Chapter 7** we describe the characterization of *S.solfataricus* ChoR, an archaeal-specific regulator for copper homeostasis. Genes encoding orthologues of ChoR are only present in archaeal genomes, where most of them are clustered with genes encoding a small putative protein (ChoC) and a putative P-type cation-transporting ATPase (ChoA). While the N-terminal part of ChoR is homologous to HTH domains of several different bacterial regulators, its C-terminal portion contains a putative metal-binding motif that resembles that of mammalian myeloproliferative disease and mental retardation (MYM) proteins. Interestingly, Choc and ChoA possess a similar motif. The *choC* and *choA* genes are transcribed as a polycistronic mRNA, which is induced when growth-inhibiting concentrations of either copper or cadmium are present in the medium, whereas *choR* is transcribed constitutively as a monocistronic mRNA. *In vitro*, ChoR binds to the *choCA* promoter and apparently functions as a copper-inactivated repressor for *choCA* transcription, because its DNA-binding affinity appears optimal in the absence of metal-ions, and lower in the presence of Cu(II). Thus, ChoA appears to be a copper and cadmium exporting ATPase, the transcription of which is induced by Cu(II) through derepression by ChoR. Copper-exporting systems in both bacteria and eukarya comprise P-type ATPases homologous to ChoA, to which copper is delivered by specific copper chaperones. In analogy, ChoCs most likely serve as copper chaperones for ChoAs encoded by archaeal *cho* gene clusters. Moreover, the N-terminal MYM-like metal binding motifs of archaeal ChoAs and ChoCs appear to be functional analogues of the N-terminal MxCxxC metal-binding motifs present in bacterial and eukaryal ATPases and their chaperones.

In this thesis, we describe the identification and characterization of transcriptional regulators of hyperthermophilic archaea. In general, transcription initiation is a multi-step process involving the formation of distinct intermediate complexes. The rate-limiting step in this sequence of events varies between different promoters, and in bacteria as well as in eukarya any of these steps can serve as the target for a transcriptional regulator. Several detailed studies, as well as the data presented in this thesis, indicate that archaeal repression can occur at the first step of transcription initiation: pre-initiation complex (PIC) assembly, either by blocking TBP/TFB binding or RNAP recruitment. Such mechanisms are most similar to the mechanism of promoter-occlusion in bacteria, although PIC assembly in bacteria involves fewer steps since the RNAP holoenzyme binds directly to the promoter. It should be noted however that the studied regulators mentioned here were selected on the basis of similarity with bacterial regulators, and they may therefore act through similar (bacterial) mechanisms. In contrast to transcriptional repression, activation usually requires a more specific interplay between (co-) regulators and general transcription factors, and it is of great interest to elucidate archaeal activating mechanisms. We have provided strong evidence for activation of archaeal transcription by a bacterial-like regulator, although we were unable to identify an anticipated co-activator in this presumed activation.

Instead of comparing archaeal regulatory mechanisms to the many mechanisms known to date, it more useful to assign archaeal mechanisms to one of the paradigms of transcriptional regulation as they occur in all three domains of life (Fig. 1). In eukarya, chromatin plays a key role in regulation of transcription. Transcription relies on the modification or remodeling of the chromatin structure, enabling the PIC to be assembled. Because of the repressed ground state by the chromatin, repression is generally not necessary. However, active repression mechanisms involving chromatin modification have been demonstrated, e.g. in silencing of developmental genes. Repression of archaeal transcription by the chromatin-associated protein acetylation lowers binding affinity (ALBA), as well as its post-translational modification, appears to be most comparable to this paradigm of regulation. Although histone-like proteins sometimes affect transcription in bacteria, they do not appear to be actively involved in regulation, and their role is believed to be merely architectural.

Activation of transcription is in many cases dependent on direct interactions between general transcription factors and RNA polymerase on one hand, and activators or co-activators on the other hand, which bind upstream of the core promoter elements. Binding sites are called upstream activating sequences (UAS) when they are in close proximity, or enhancers when they are located relatively far upstream of the core promoter elements. In

the case of enhancers, the above-mentioned interactions require DNA looping, which in bacteria is sometimes facilitated by the binding of a DNA-bending protein like integration host factor (IHF). Interactions are not always direct, but may be mediated by other proteins or co-activators. Whereas these proteins can be single proteins, in eukarya this involves a large multi-subunit protein complex called mediator, the size and composition of which is highly variable between different eukarya. Likewise, the presumed activation by LysM in archaea appears to be mediated by a yet unidentified protein (complex).

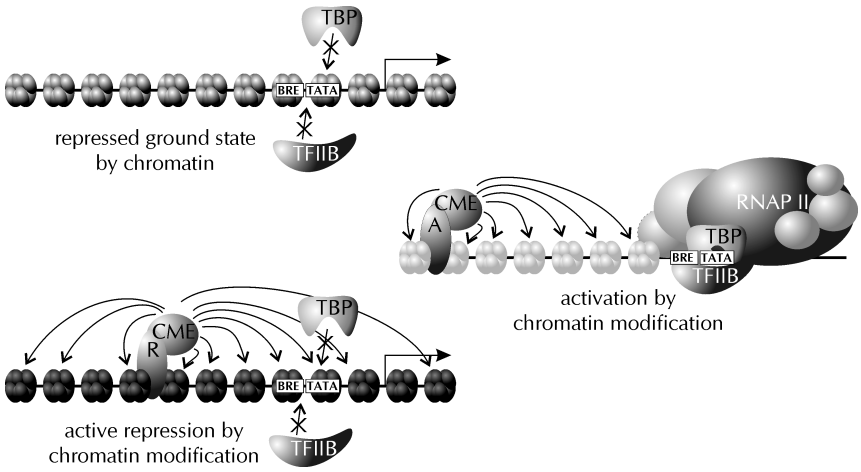
The paradigm for repression in bacteria is called promoter-occlusion. This involves the binding of a repressor to a site overlapping (one of the) core promoter elements. In contrast to eukarya, which generally do not use this mechanism, archaea have now been shown to utilize a variety of repression mechanisms that rely on promoter-occlusion. In contrast to the one-step process of PIC-formation in bacteria, PIC-formation in archaea is a multi-step process. Accordingly, the promoter-occlusion occurs at two levels: TBP/TFB-binding and RNAP recruitment.

Although examples of archaeal transcriptional regulation mechanisms are just beginning to emerge, it seems that they are modifications of the mechanisms encountered in both eukarya and bacteria. It is anticipated that as with many cellular processes of archaea, the complete range of archaeal transcriptional regulatory mechanisms may be a mosaic of bacterial and eukaryal features.

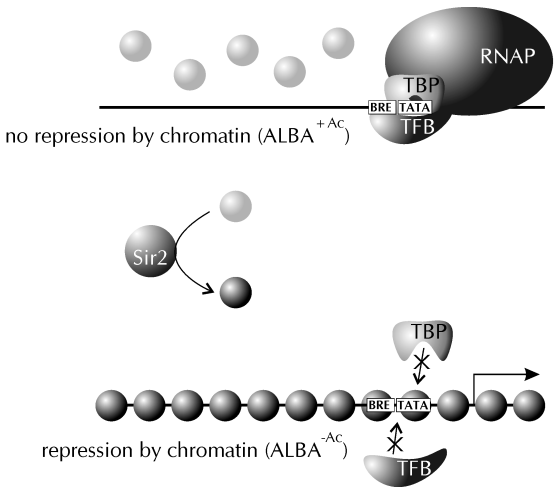
**Figure 1.** (next page) Paradigms for transcriptional regulation in eukarya, archaea, and bacteria. A, activator; R, repressor; CME, chromatin-modifying enzyme; C, co-activator. See text for explanation.

regulation by chromatin structure

eukarya



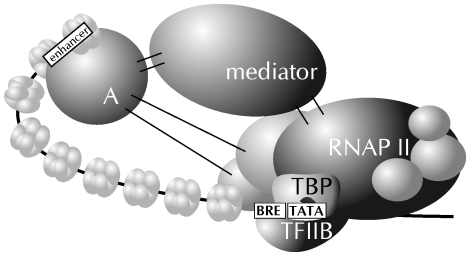

archaea

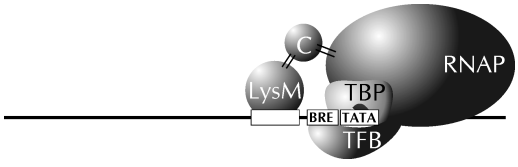


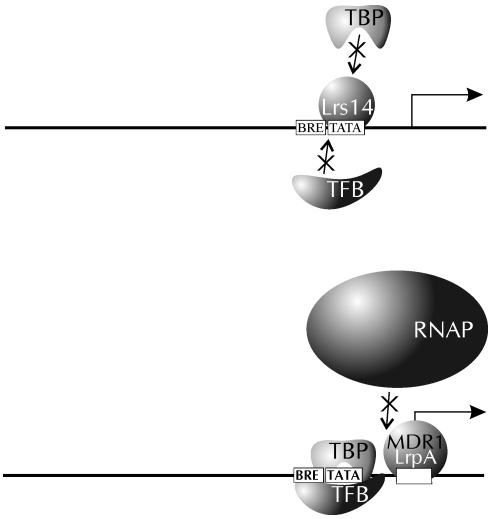

bacteria

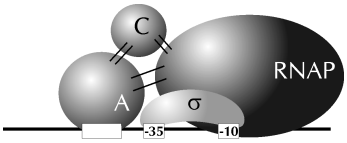
activation through (in)direct contacts

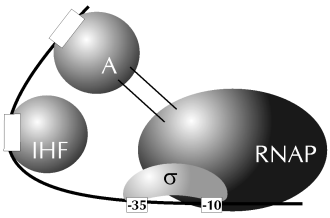
repression by promoter-occlusion

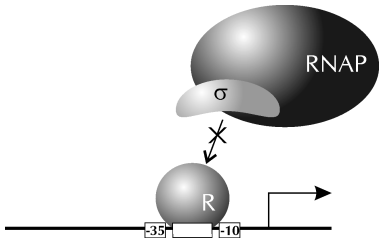


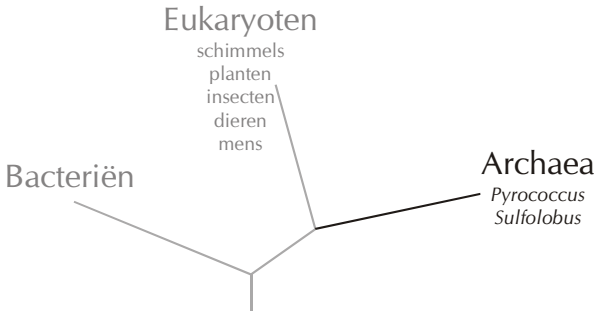



# *Hoofdstuk 9*

Nederlandse samenvatting

## Eukaryoten, bacteriën en archaea: de drie domeinen

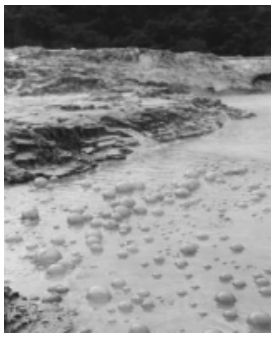
Alle levende organismen kunnen verdeeld worden in drie groepen of "domeinen": eukaryoten, bacteriën en archaea. Het domein van de eukaryoten bevat schimmels, planten, insecten, vogels en zoogdieren. Veel van de eukaryoten zijn complexe meercellige organismen: de mens bestaat bijvoorbeeld uit miljarden cellen. Bacteriën kennen we vooral als ziekteverwekkers, maar er zijn minstens zoveel onschadelijke bacteriën die kunnen leven op allerlei plaatsen. Bacteriën zijn ééncellige, voor het blote oog onzichtbare organismen. Archaea lijken erg veel op bacteriën. Ook dit zijn kleine, ééncellige organismen. Toch verschillen ze zoveel van bacteriën dat ze in een apart domein worden ingedeeld.




## *Pyrococcus furiosus* en *Sulfolobus solfataricus*

Veel archaea zijn extremofielen; zij leven op plaatsen waarvan je zou denken dat er geen leven mogelijk is. Zo zijn er archaea die houden van een extreem zoute omgeving, extreme druk, pH of temperatuur. Wanneer archaea zich optimaal kunnen delen bij een temperatuur van boven de 80°C worden ze hyperthermofiel genoemd. In deze categorie vallen de archaea die onderzocht werden in dit proefschrift. De twee hyperthermofiele organismen uit dit proefschrift zijn *Pyrococcus furiosus* en *Sulfolobus solfataricus*. *Pyrococcus furiosus* betekent letterlijk "ziedende vuurbal" omdat hij rond van vorm is en zich deelt bij 100°C. Ziedend duidt in dit geval op een erg snelle celdeling. *P.furiosus* is geïsoleerd bij het eiland Vulcano vlakbij Napels, Italië. Verwante organismen van het genus *Pyrococcus* leven op verschillende plaatsen van de wereld in de buurt van black smokers. Dit zijn geisers die zich op grote diepte op de zeebodem bevinden. *Sulfolobus solfataricus* werd geïsoleerd in een vulkanisch gebied bij Napels dat Solfatara heet, uit zwavelrijke poelen die extreem zuur zijn (pH 3) en een temperatuur hebben van 80°C. Verwante organismen van het genus

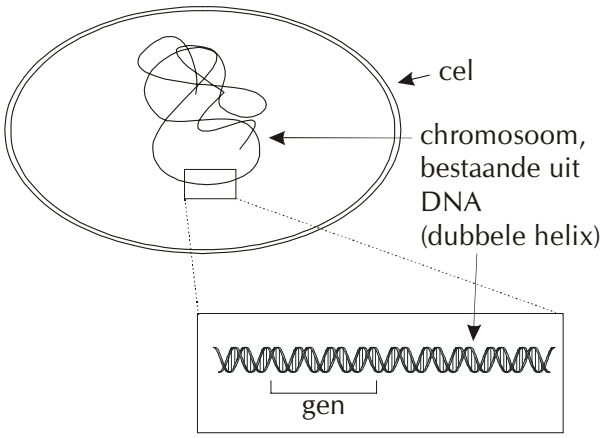

Black smoker


Solfatara

*Sulfolobus* zijn ook te vinden in bijvoorbeeld Yellowstone National Park (USA), waar zulke poelen ook aanwezig zijn.

## DNA

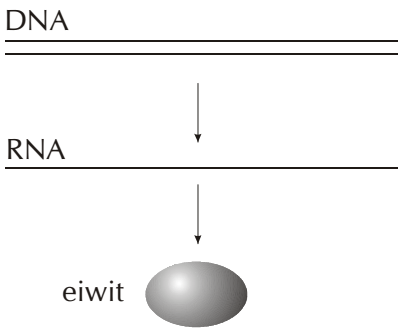
Hoewel er belangrijke verschillen zijn tussen eukaryoten, bacteriën en archaea, zijn er ook overeenkomsten. Alle levende cellen bevatten DNA waarin de genetische informatie ligt opgeslagen. Deze genetische informatie is de blauwdruk van de cel en bepaalt alle eigenschappen die een cel heeft. DNA bestaat uit slechts vier verschillende bouwstenen (A, C, G, T). Deze bouwstenen –de basen– kunnen in elke willekeurige volgorde achter elkaar voorkomen, zodat ze samen een lange streng vormen. DNA bestaat uit twee strengen die bij elkaar horen, om elkaar gedraaid zitten en een dubbele helix vormen. Zo'n paar van om elkaar gedraaide strengen noemt men een chromosoom. Archaeële en bacteriële cellen hebben één chromosoom, terwijl menselijke cellen er bijvoorbeeld 46 hebben (2 x 22 + x+y). Ook de lengte van chromosomen varieert nogal: het chromosoom van *S.solfataricus* bestaat uit 3 miljoen basen, terwijl alle menselijke chromosomen samen meer dan 3 miljard basen tellen. Dit zijn enorme aantallen basen: het chromosoom van *S.solfataricus* telt al gauw 700 pagina's tekst aan A's, C's, G's en T's. Chromosomen zijn ingedeeld in functionele segmenten, die genen worden genoemd. Genen worden gekenmerkt door een start– en een stopsignaal. Een chromosoom is dus te vergelijken met een lang stuk tekst dat alleen de letters A, C, G, of T bevat. Een gen is een zin uit deze tekst die herkend kan worden aan een hoofdletter aan het begin en een punt aan het einde.




## Transcriptie en translatie

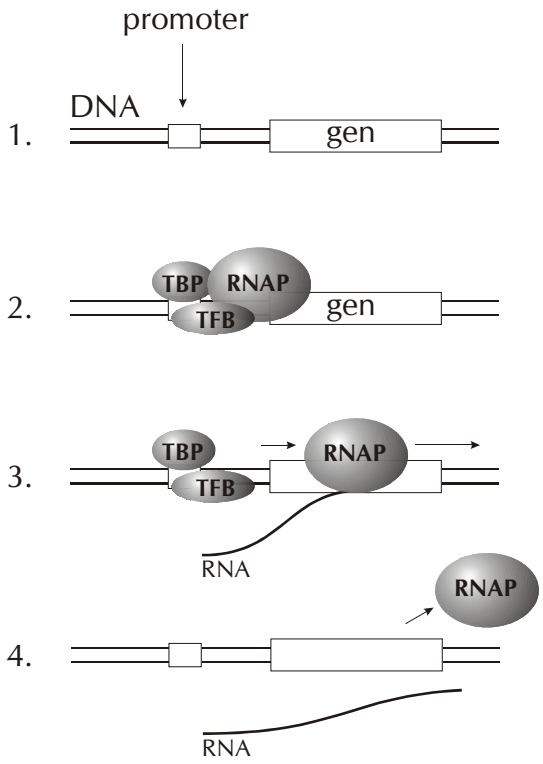
Wat doet een cel met zijn DNA? Om gebruik te maken van zijn genen worden deze "overgeschreven" op een tijdelijke informatiedrager die uit RNA bestaat. RNA bevat nagenoeg dezelfde bouwstenen als DNA (A, C, G en U) maar bestaat slechts uit één streng. Dit proces van overschrijven van DNA naar RNA heet transcriptie. De basenvolgorde van het RNA wordt vervolgens gedecodeerd waardoor er een eiwit gemaakt kan worden. Dit

proces heet translatie. De identiteit en de eigenschappen van het eiwit worden bepaald door de basenvolgorde van het RNA, en omdat RNA overgeschreven is van het DNA bepaalt de basenvolgorde van het gen op het DNA dus hoe betreffende gecodeerde eiwit eruit ziet en wat de functie is. Eiwitten kunnen verschillende functies hebben. Zo hebben sommige eiwitten een structurele functie. Dit is bijvoorbeeld het geval bij eiwitten uit de celmembraan. Andere eiwitten hebben een katalytische functie en helpen bij de omzetting van stof A naar stof B. In dit geval noem je het een enzym. Enzymen helpen bijvoorbeeld bij de voedselvertering door grotere moleculen in kleinere, bruikbare stukken te hakken. Op deze manier zijn genen, bestaande uit DNA, verantwoordelijk voor een bepaalde functie die een cel nodig heeft om te kunnen leven.




## Regulatie van transcriptie

Nu kunnen er voor een cel allerlei redenen zijn om een gen niet *altijd* om te zetten naar RNA en eiwit. Transcriptie en translatie kosten de cel bijvoorbeeld energie, dus alleen als een eiwit echt nodig is wordt het betreffende gen op het DNA overgeschreven naar RNA en vertaald naar een eiwit. Neem bijvoorbeeld het gen dat codeert voor een enzym dat lactose omzet in het voor de cel bruikbare glucose. Er zou veel energie verspild worden als dit enzym gemaakt zou worden wanneer er geen lactose aanwezig is. Daarom worden processen als transcriptie en translatie gereguleerd. Alleen wanneer het nodig is worden genen overgeschreven naar RNA en vertaald naar een eiwit.




Laten we transcriptie, oftewel het overschrijven van een gen bestaande uit DNA naar RNA, eens bekijken. Zoals gezegd wordt een gen gekenmerkt door een

start– en een stopteken. Archaeële genen zijn gemiddeld zo'n 1000 tot 1500 basen lang. Voorafgaand aan het startteken ligt een promoter. Dit is een basenvolgorde die herkend wordt als startplaats voor transcriptie. In archaea bevatten promoters een reeks basen waarin de volgorde "TATA" voorkomt. Deze TATA-box wordt specifiek herkend door een eiwit dat TATA-binding protein heet. TATA-binding protein kan zo'n TATA-box herkennen en er vervolgens aan binden. Wanneer dat gebeurt komt er een tweede eiwit, transcriptie factor B, dat hier op zijn beurt weer aan bindt. Als laatste bindt hieraan een groot eiwitcomplex dat RNA polymerase heet. RNA polymerase is nu in staat om, startende vanaf de promoter, over het DNA heen te schuiven, de basenvolgorde te lezen en te kopiëren in RNA met dezelfde basenvolgorde. Wanneer dit gebeurt vindt er dus transcriptie plaats, en wordt er een streng RNA gemaakt waarop een gen ligt. Wanneer RNA polymerase het stopteken van het gen gepasseerd heeft wordt transcriptie meestal beëindigd doordat RNA polymerase het DNA loslaat. In bacteriën en archaea is het echter mogelijk dat RNA polymerase meerdere genen in één keer kopieert. In dit geval bevat één RNA streng dus meerdere genen, en spreekt men van een operon.

Regulatie van transcriptie vindt meestal plaats bij de promoter, waar TATA-binding protein, transcriptie factor B en RNA polymerase binden. Door de binding van deze eiwitten aan het DNA positief of negatief te beïnvloeden kan transcriptie gestimuleerd respectievelijk geremd worden. Dit wordt bewerkstelligd door eiwitten die regulatoren genoemd worden. Vaak zijn dit DNA-bindende eiwitten, die repressors heten als ze transcriptie remmen en activators heten als ze transcriptie stimuleren. Regulatie van transcriptie is van groot belang voor elke levende cel. Wanneer er geen adequate regulatie plaatsvindt lopen cellulaire processen uit de rails. Een extreem voorbeeld hiervan is het mislopen van regulatie van transcriptie in menselijke cellen, waardoor kanker kan ontstaan.

In dit proefschrift hebben we de regulatie van transcriptie in *P.furiosus* en *S.solfataricus* bestudeerd. In tegenstelling tot regulatie van transcriptie in bacteriën en eukaryoten is er slechts weinig bekend over regulatie van transcriptie in deze hyperthermofiele archaea. Dit komt omdat het basale mechanisme van transcriptie in archaea nog niet zo lang bekend is. Vreemd genoeg lijkt dit mechanisme meer op dat van eukaryoten dan op dat van bacteriën (zie Hoofdstuk. 2, Figuur 2), iets wat je niet zou verwachten omdat archaea wat betreft cellulaire organisatie juist meer op bacteriën lijken. Het doel van dit onderzoek is daarom vooral fundamenteel wetenschappelijk, om te begrijpen hoe een belangrijk proces als transcriptie regulatie in archaea werkt. Aan de andere kant kan de verkregen kennis in de toekomst nuttig zijn wanneer hyperthermofiele archaea ingezet worden in

biotechnologische processen zoals de productie van enzymen met behulp van *P.furiosus* of *S.solfataricus*.

## De technieken

DNA, RNA en eiwitten zijn te klein om met een microscoop te bestuderen. Om een idee te geven van hun afmetingen: een DNA fragment bestaande uit 1000 basen is slechts 0,34 duizendste millimeter lang. Om regulatie van transcriptie te bestuderen moet je dus gebruik maken van moleculaire technieken. Een veel gebruikte techniek in dit proefschrift is gel-elektroforese. DNA, RNA of eiwitten kunnen hiermee bijvoorbeeld op grootte gescheiden worden door ze in een gel te brengen en hier vervolgens een elektrische spanning op aan te brengen. Omdat DNA, RNA en eiwitten negatief geladen zijn zullen ze naar de positieve pool gaan bewegen. Een korte DNA streng kan dat sneller dan een lange, en na verloop van tijd zullen fragmenten van verschillende lengte zich op verschillende plaatsen in de gel bevinden. Dit zijn de "bandjes" die je ziet op veel van de figuren in dit proefschrift. De gebruikte hoeveelheden DNA, RNA en eiwit zijn erg klein, variërend van minder dan een miljardste (ng) tot een miljoenste gram (μg). Daarom is het vaak nodig DNA zichtbaar te maken door het radioactief te labelen.

Aangezien de te bestuderen eiwitten uit *P.furiosus* en *S.solfataricus* maar in kleine hoeveelheden aanwezig zijn in deze organismen hebben we in dit proefschrift gebruik gemaakt van recombinant DNA technieken. Hierbij maken we gebruik van *Escherichia coli*, een bacterie die hiervoor gebruikt wordt in vrijwel alle laboratoria. Het gen dat codeert voor het eiwit van interesse wordt uit bijvoorbeeld *S.solfataricus* geïsoleerd en ingebracht in *E.coli*. Deze zal het gecodeerde eiwit in grote hoeveelheden gaan maken. Transcriptie en translatie zijn immers processen die in elk levend organisme plaatsvinden en de DNA code is universeel. Op deze manier is het mogelijk een grote hoeveelheid (milligrammen) eiwit te krijgen die na zuivering gebruikt kan worden in verdere experimenten.

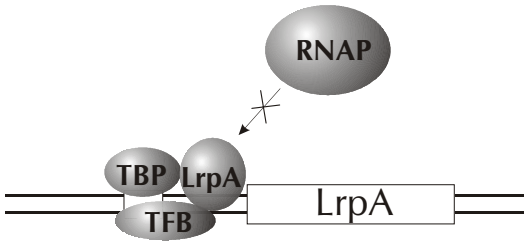
## Resultaten

In Hoofdstuk 2 hebben we een gen gevonden op het chromosoom van *S.solfataricus* waarvan de functie niet bekend was. Echter, een gen wat hierop lijkt is ook te vinden in eukaryote organismen waar het essentieel is voor transcriptie. Het gen van *S.solfataricus* codeerde voor het eiwit TFE (transcriptie factor E). Het TFE eiwit bleek transcriptie in *S.solfataricus* in sommige gevallen te kunnen stimuleren. Stimulatie door TFE vond alleen plaats wanneer transcriptie startte van een promoter die een "slechte" TATA-box had. Dit is een TATA-box waarvan de basenvolgorde net iets afwijkt van de normale TATA-box

zodat hij niet goed herkend wordt door TBP. Het is dan ook waarschijnlijk dat TFE transcriptie van dit soort promoters in *S.solfataricus* stimuleert door de herkenning en binding van TBP aan de TATA-box te verbeteren.

Hoofdstuk 3 is een literatuurstudie over een eiwitfamilie die de Lrp-familie heet (leucine-responsive regulatory protein). Lrp eiwitten komen voor in bacteriën en archaea. In bacteriën zijn deze eiwitten betrokken bij regulatie van transcriptie, maar hun rol in archaea was onbekend bij aanvang van ons onderzoek.

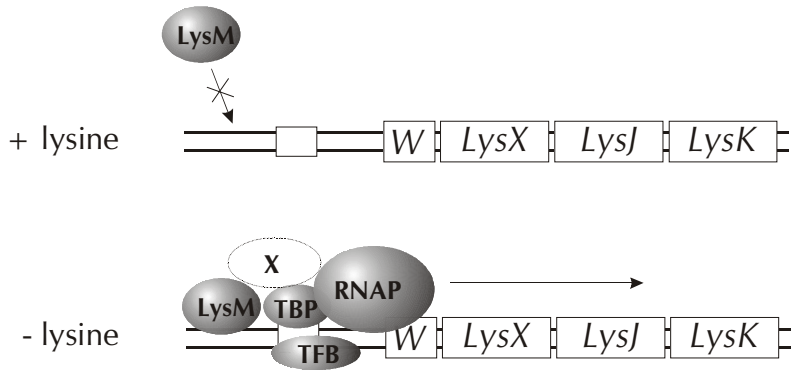
In Hoofdstuk 4 namen we voor het eerst zo'n Lrp eiwit van *P.furiosus* (LrpA) onder de loep. Ons onderzoek wees uit dat LrpA een DNA-bindend eiwit was dat kon binden aan een specifieke basenvolgorde in de promoter van het gen coderend voor LrpA. Wanneer LrpA daaraan bond werd transcriptie van het LrpA gen geremd. LrpA fungeerde dus als een repressor en reguleerde daarmee zijn eigen transcriptie. De plaats waar LrpA bond aan zijn promoter suggereerde dat transcriptieremming berustte op het uitsluiten van RNA polymerase binding, terwijl binding van TATA-binding protein en transcriptiefactor B ongehinderd bleef. De negatieve autoregulatie door LrpA heeft waarschijnlijk tot doel een constante LrpA concentratie in de cel te handhaven en werkt dus als een soort thermostaat voor de LrpA promoter. Het is aannemelijk dat het LrpA eiwit er niet slechts is om zijn eigen transcriptie te reguleren en we denken dan ook dat er in *P.furiosus* nog een andere promoter is waar LrpA aan bindt. Hoewel we hiernaar hebben gezocht hebben we zo'n promoter helaas niet kunnen vinden.




In Hoofdstuk 5 hebben we de structuur van het LrpA eiwit opgelost zodat het mogelijk is het eiwit te bekijken. Zoals gezegd zijn DNA, RNA en eiwitten erg klein en kun je ze met de microscoop niet zien. Toch is het voor het begrijpen van de functie van LrpA en andere eiwitten uit de Lrp familie noodzakelijk om te weten hoe zo'n Lrp eiwit eruit ziet. Daarom hebben we gebruik gemaakt van kristallisatie van het LrpA eiwit. Onder de juiste condities kristalliseren eiwitten zoals ook zoutkristallen ontstaan wanneer je zeewater indampt. Wanneer je röntgenstralen op een eiwitkristal loslaat zullen deze verspreid worden in een bepaald patroon (röntgendiffractie). Aan de hand van dit patroon kan dan de 3D structuur van het eiwit berekend worden en dus zichtbaar gemaakt worden. De structuur van LrpA was de eerste structuur van een eiwit uit de Lrp familie en het is aannemelijk dat andere Lrp eiwitten uit deze familie er ongeveer hetzelfde uitzien. De structuur van LrpA liet onder meer zien dat LrpA twee delen heeft waarvan duidelijk is dat één deel kan binden

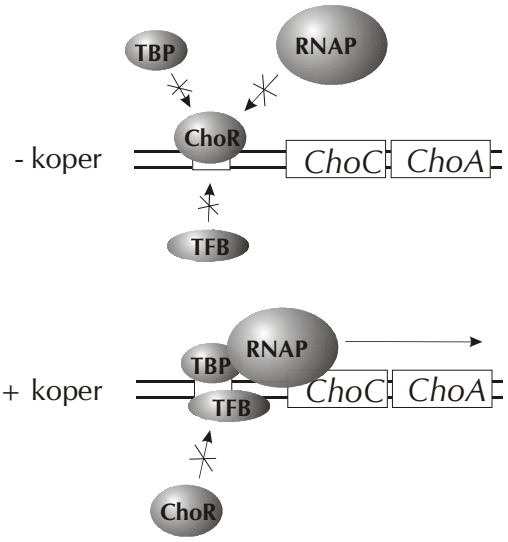
aan DNA terwijl het andere deel fungeert als sensor voor een nog onbekende stof (zie Hoofdstuk 5, Figuur 1).

In Hoofdstuk 6 hebben we LysM, een ander eiwit uit de Lrp familie, bestudeerd. LysM is afkomstig van *S.solfataricus*. Het gen coderend voor LysM ligt op het DNA tussen andere genen waarvan we vermoedden dat ze betrokken waren bij de aanmaak (biosynthese) van het aminozuur lysine. We noemden deze genen samen dan ook het Lys-gencluster. Het aminozuur lysine is één van de 20 verschillende aminozuren die de bouwstenen zijn van eiwitten. Lysine is dus onmisbaar voor elke cel. Om te verifiëren of het Lys-gencluster inderdaad een rol speelt bij lysine biosynthese hebben we de transcriptie van de genen uit dit cluster bestudeerd in *S.solfataricus* cellen die groeiden in aan- of afwezigheid van lysine. Eén van de promoters van het gencluster bleek inderdaad gereguleerd door lysine: in afwezigheid van lysine vond er wel transcriptie plaats en in aanwezigheid van lysine niet. Deze promoter bleek de startplaats te zijn voor transcriptie van vier van de zeven Lys genen uit het cluster. Verder onderzoek wees uit dat het LysM eiwit deze promoter kon herkennen en binden. Bovendien vonden we dat lysine de binding van LysM aan de promoter zwakker maakte. Onze hypothese is daarom dat LysM een sensor is voor lysine en een activator is voor deze promoter: in afwezigheid van lysine vindt er transcriptie plaats van de Lys genen om lysine biosynthese mogelijk te maken. Om dit te bewerkstelligen bindt LysM en stimuleert transcriptie. In afwezigheid van lysine hoeft *S.solfataricus* zelf geen lysine te maken. Er vindt geen transcriptie plaats omdat LysM niet kan binden aan de promoter en transcriptie dus niet kan stimuleren. Helaas was het niet mogelijk om dit model direct te testen. We nemen daarom aan dat er nog een ander (onbekend) eiwit nodig is bij de stimulatie van transcriptie.




In de voorgaande hoofdstukken hebben we eiwitten uit de Lrp familie bestudeerd. Dit zijn eiwitten die in bacteriën en archaea voorkomen. In Hoofdstuk 7 hebben we het eiwit ChoR bestudeerd dat alleen in archaea voorkomt. Het gen coderend voor ChoR ligt naast twee genen (ChoC en ChoA) coderend voor eiwitten die betrokken zijn bij het handhaven van een constante concentratie koper in *S.solfataricus* cellen (Cho staat voor copper

homeostasis). Elke cel heeft koper nodig, maar slechts een kleine hoeveelheid. Wanneer er te veel koper aanwezig is ontstaan er ernstige problemen. Als er in bijvoorbeeld menselijke cellen door een defect te veel koper aanwezig is ontstaan er ernstige ziekten zoals de ziekten van Menkes en Wilson. ChoA uit *S.solfataricus* lijkt erg op de eiwitten die defect zijn bij mensen met deze ziekten, en functioneert als transport-eiwit dat koper de cel uitwerkt. ChoC is waarschijnlijk een nog niet eerder beschreven chaperonne-eiwit dat het koper in de cel vasthoudt en begeleidt zodat het geen schade aanricht. Wij hebben de Cho genen gebruikt om regulatie van transcriptie in *S.solfataricus* te bestuderen. We vonden dat transcriptie van de ChoC en ChoA genen startte vanaf één promoter, maar dat er normaal gesproken geen transcriptie was vanaf deze promoter. Alleen wanneer er een hoge koper concentratie aanwezig was vond er transcriptie van ChoC en ChoA plaats vanaf deze promoter. We vonden dat ChoR functioneerde als negatieve regulator voor deze promoter: ChoR bond namelijk alleen aan deze promoter in de afwezigheid van koper. In de aanwezigheid van koper kon ChoR niet binden aan de promoter.




## Conclusie

Hoewel archaea veel overeenkomsten hebben met bacteriën lijkt hun transcriptiesysteem meer op dat van eukaryoten. Veel archaeële regulatoren van transcriptie lijken echter wél op die van bacteriën en voorafgaand aan ons onderzoek was dan ook niet duidelijk hoe deze bacterieel-achtige regulatoren dienst doen in een eukaryoot-achtig transcriptiesysteem. In ons onderzoek hebben we een viertal eiwitten uit archaea onderzocht die transcriptie kunnen beïnvloeden of reguleren. TFE stimuleert transcriptie van promoters met slechte TATA-boxen. LrpA en LysM behoren tot de Lrp-familie van eiwitten die ook in bacteriën voorkomen. Uit ons onderzoek blijkt dat deze eiwitten dezelfde soort functies vervullen als in bacteriën. Ze kunnen transcriptie remmen (LrpA) maar lijken ook in staat transcriptie te stimuleren (LysM). Bovendien bleek LysM een sensor te zijn voor het aminozuur lysine. Bacteriële Lrp-eiwitten dienen ook als sensoren voor aminozuren. Daarnaast hebben we LrpA gekristalliseerd en de 3D structuur van LrpA opgelost. Het vierde eiwit (ChoR) is specifiek voor archaea en komt in bacteriën niet voor.

ChoR is betrokken bij handhaving van een aanvaardbare koperconcentratie in *S.solfataricus* en remt transcriptie van de ChoC en ChoA genen.

De mechanismen van regulatie die wij bestudeerd hebben lijken vergelijkbaar te zijn met bacteriële mechanismen, waarbij remming van transcriptie bijvoorbeeld wordt bewerkstelligd doordat het betrokken regulator-eiwit binding van TATA-binding protein, transcriptiefactor B of RNA polymerase aan de promoter uitsluit.

## *Curriculum vitae*

Arjen Brinkman werd geboren in Amerongen op 7 januari 1973. In dezelfde plaats groeide hij op en doorliep er de kleuterschool en de lagere school. Van 1985 tot 1992 ging hij in Doorn naar het Revius Lyceum, waar hij eerst zijn HAVO diploma en later zijn VWO diploma behaalde. Het jaar daarop volgde hij aan het Koninklijk Conservatorium in Den Haag de opleiding tot uitvoerend jazz-gitarist. In 1993 echter besloot hij het roer om te gooien en schreef hij zich in bij het Hoger Laboratorium Onderwijs van de Hogeschool Utrecht. Hij volgde de medisch-biotechnologische richting en in het laatste jaar van deze opleiding liep hij stage in Salt Lake City, VS, waar hij onder prof. David Low werkte aan regulatie van transcriptie in *Escherichia coli*. In 1997 behaalde hij zijn HLO diploma en werkte hij aansluitend als onderzoeksanalist bij de vakgroep Medische Microbiologie van de Vrije Universiteit in Amsterdam. Hier werkte hij onder dr. Hans Kusters en prof. dr. Christina Vandenbroucke-Grauls aan een project waarin mechanismen van antibioticaresistentie in de maagbacterie *Helicobacter pylori* bestudeerd werden. In 1998 besloot hij Onderzoeker in Opleiding te worden bij het Laboratorium voor Microbiologie van Wageningen Universiteit, onder prof. dr. Willem M. de Vos en dr. John van der Oost. Hier bestudeerde hij de regulatie van transcriptie in hyperthermofiele archaea, waarvan de resultaten beschreven staan in dit proefschrift. Vanaf april 2002 werkt hij als Postdoc bij de Vakgroep Fysiologische Chemie van de Universiteit Utrecht, waar hij het mechanisme van Adenovirus DNA-replicatie bestudeert.

## *List of publications*


Debets-Ossenkopp, Y.J., Brinkman, A.B., Kuipers, E.J., Vandenbroucke -Grauls, C.M., Kusters, J.G., 1998, Antimicrobial Agents and Chemotherap y, 42(10): 2749-2751, Explaining the bias in the 23S rRNA gene mutations associated with clarithromycin resistance in clinical isolates of *Helicobacter pylori.*

Brinkman, A.B., Dahlke, I., Tuininga, J.E., Lammers, T., Dumay, V., de Heus, E., Lebbink, J.H.G ., Thomm, M., de Vos, W.M., van der Oost, J., 2000, Journal of Biological Chemistry 275(49): 38160 -38169, An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated.

Sedelnikova, S.E., Smits, S.H.J., Leonard, P .M., Brinkman, A.B., van der Oost, J. Rafferty, J.B. and Rice, D.W., (2001), Acta Cryst. D57: 886 -888, Crystallisation of an Lrp -like regulatory protein from the hyperthermophile *Pyrococcus furiosus.*

Leonard, P.M., Smits S.H.J., Brinkman, A.B., de Vos, W .M., van der Oost, J., Rice, D.W., Rafferty, J.B., 2001, EMBO Journal 20(5): 990 -997, Crystal structure of the Lrp -like transcriptional regulator from the archaeon *Pyrococcus furiosus.*

Bell, S.D., Brinkman, A.B., van der Oost, J., Jackson, S.P., (2001), EMBO Reports 2(2): 133 -138, The archaeal TFIIEα homologue facilitates transcription initiation by enhancing TATA -box recognition.

Brinkman, A.B., Bell, S.D., Lebbink, R.J., de Vos, W.M., van der Oost, J., 2002, Journal of Biological Chemistry, *in press*, The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability.

Ettema, T.J.G., Brinkman, A.B., Leonard, P.M., Tani, T.H., Rafferty, J.B., van der Oost, J. ,2002, Journal of Biological Chemistry, *in press*, A novel ligand -binding domain (RAM) involved in allosteric regulation of amino acid metabolism in prokaryotes.