The protein ORF80 from the acidophilic and thermophilic archaeon *Sulfolobus islandicus* binds highly site-specifically to double-stranded DNA and represents a novel type of basic leucine zipper protein

Georg Lipps*, Pablo Ibanez, Thomas Stroessenreuther, Katya Hekimian and Gerhard Krauss

University Bayreuth, Biochemistry II, Universitätsstrasse 30, 95447 Bayreuth, Germany

Received September 6, 2001; Revised and Accepted October 22, 2001

**ABSTRACT**

The cryptic high copy number plasmid pRN1 from the thermophilic and acidophilic crenarchaeote *Sulfolobus islandicus* shares three conserved open reading frames with other *S.islandicus* plasmids. One of the open reading frames, namely orf80, encodes a 9.5 kDa protein that has no homology to any characterised protein. Recombinant ORF80 purified from *Escherichia coli* binds to double-stranded DNA in a sequence-specific manner as suggested by EMSA experiments and DNase I footprints. Two highly symmetrical binding sites separated by 60 bp were found upstream of the *orf80* gene. Both binding sites contain two TTAA motifs as well as other conserved bases. Fluorescence measurements show that short duplex DNAs derived from a single binding site sequence are bound with submicro-molar affinity and moderate cooperativity by ORF80. On DNA fragments carrying both binding sites, a rather large protein–DNA complex is formed in a highly cooperative manner. ORF80 contains an N-terminal leucine zipper motif and a highly basic domain at its C-terminus. Compared to all known basic leucine zipper proteins, the order of the domains is reversed in ORF80. ORF80 may therefore constitute a new subclass of basic leucine zipper DNA-binding proteins.

**INTRODUCTION**

Plasmid pRN1 from the acidophilic and thermophilic archaeon *Sulfolobus islandicus* was first described in 1994 by Zillig et al. (1). pRN1 belongs to the archaeal plasmid family pRN (2), which now includes five members: plasmids pRN2 (3), pSSVx (4) and pHEN7 (2) have been isolated from *S.islandicus* strains; another member of this family, pDL10, is found in the crenarchaeote *Acidimius ambivalens*.

The replication mode of the pRN plasmid family has not yet been investigated in detail. Indirect evidence exists that pDL10 replicates via a rolling circle mechanism (5). The pRN plasmids share a core of three highly conserved open reading frames. The shortest gene, *orf56* in pRN1, has low homology to plasmid copy control proteins of several eubacterial plasmids (3). We could show that purified recombinant ORF56 binds with high affinity to an imperfect inverted repeat within its promoter (6), suggesting that ORF56 may inhibit its own expression and expression of the overlapping gene *orf904*, which does not seem to have a promoter of its own. Another short open reading frame, *orf80* in pRN1, displays no homology to any characterised protein. It was tentatively named plasmid regulatory protein (plr) (5). The third conserved open reading frame codes for a large protein of about 900 amino acids named ORF904. One domain of ORF904 shows homology to the helicase domain of the bacteriophage P4 primase and ORF904 is possibly directly involved in plasmid replication. Although this plasmid family has been very well analysed at the sequence level, information on the biochemical properties and functions of the gene products is still lacking. The plasmid family and especially the relatively small plasmids pRN1 and pRN2 are an attractive backbone for the construction of a high copy number *Sulfolobus–Escherichia coli* shuttle vector.

In this report we focus on the product of the *orf80* gene from pRN1. Homologues of the gene *orf80* are also found in the conjugative plasmids pNOB8 (7) and pING1 (8) and within the genome of *Sulfolobus solfataricus*. We expressed *orf80* in *E.coli* and characterised the purified recombinant protein. ORF80 binds specifically to two sites located upstream of its own gene. The two binding sites have distinct DNA elements in common and are of palindromic structure. Interestingly, ORF80 could constitute a new subclass of basic leucine zipper DNA-binding proteins, where the N-terminal leucine zipper is followed by a C-terminal basic domain.

*To whom correspondence should be addressed. Tel: +49 921 55 2419; Fax: +49 921 55 2432; Email: georg.lipps@uni-bayreuth.de*

Present address:
Pablo Ibanez, INSERM U289, Hôpital de la Pitie Salpêtrière, 47 Boulevard de l’Hôpital, 75013 Paris, France
MATERIALS AND METHODS

Cloning and expression of orf80

All cloning steps were performed in *E. coli* XL1-Blue. The orf80 gene was amplified from plasmid pUC18-pRN1 (a gift from David Faguy, Dalhousie University, Halifax, Canada) by PCR with forward primer 5′-AAGAATCTCAGGTATCTCATATGAGTCATGGGATC-CCATATTGAGTA and reverse primer 5′-GCGGTACGAGTCTCGTGACTCTCGAGTCCAAAC. The PCR product was cut with *Ncol* and *XhoI* (New England Biolabs) and ligated into the expression vector pET28 (Novagen) cut with the same enzymes. The identity of the insert of the resulting plasmid (pET28-orf80) was confirmed by DNA sequencing. The final construct was transformed into *E. coli* BL21(DE3) cells, which were used for overexpression. Aliquots of 4 ml of an overnight culture were inoculated into 1 l of LB medium supplemented with 50 μg/ml kanamycin and grown at 37°C. After 3 h, 1 mM IPTG was added and the culture was fermented for a further 3 h. The cells were collected by centrifugation and kept frozen at −70°C until use.

Purification of ORF80

The frozen cells were resuspended in 20 ml of lysis buffer (50 mM sodium phosphate, 1 mM EDTA, pH 8.0). Lysozyme (100 μg/ml) and Triton X-100 (0.1%) were added and the cells were incubated for 15 min at 4°C. After sonication the lysate was clarified by centrifugation for 30 min at 4°C and 40,000 g. The pH of the supernatant was adjusted to pH 6.0 and then loaded onto a fractogel EMD-DEAE column (16 ml). The column was developed with a linear salt gradient from 0.2 to 2 M Na₂SO₄ in 25 mM citrate, pH 6.0. ORF80 eluted at ~1 M Na₂SO₄. Protein-containing fractions were pooled and diluted to 0.3 M Na₂SO₄ with double distilled water. The pool was then loaded onto a 1 ml hydroxypatite column (Macroprep; Bio-Rad), which was developed with a linear salt gradient from 0.2 to 2 M NaCl in 50 mM urea, 5 mM potassium phosphate, pH 7.0. ORF80 eluted as a sharp peak at 1 M NaCl. ORF80-containing fractions were pooled and dialysed against 30% glycerol in 25 mM sodium phosphate, pH 8.0. The yield from this purification was highly variable and rather low, with ~0.3 mg ORF80 per 1 expression culture, as the majority of ORF80 aggregated during purification and had to be removed by centrifugation.

During the course of the study we learned that ORF80 could also be purified using an alternative approach. Cell extracts were precipitated with 20% saturated ammonium sulphate. The precipitated protein was washed with water and the water-insoluble ORF80 was solubilised in 6 M urea, 300 mM NaCl, 50 mM sodium phosphate, pH 7.0, and purified on an EMD-50 column in the presence of 6 M urea. ORF80 purified by this protocol was indistinguishable in electrophoretic mobility shift assays (EMSA) from ORF80 purified by the former protocol. The alternative purification scheme allowed us to obtain and store ORF80 at concentrations >50 μM, which we needed to perform the CD and fluorescence measurements.

The concentration of the purified protein was calculated using the theoretical extinction coefficient ε = 11 744 M⁻¹ cm⁻¹.

EMSA

First, the DNA-binding activity of ORF80 was assayed with restriction fragments obtained from pUC18-pRN1, which contains the complete sequence of pRN1. ORF80 was incubated for 2 min with the restriction digests in 20 mM Tris–acetate, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9. DNA fragments were then separated on 1% agarose gels (45 mM Tris–borate, 0.5 mM EDTA, pH 8.3) and stained with ethidium bromide.

After identification of the binding site by footprinting experiments, 32P-labelled oligodeoxynucleotides encompassing the ORF80 binding site were hybridised with a complementary oligodeoxynucleotide and used as probes in EMSA (Table 1).

| BS-36bp | AAACATTTAATAGACCTTAACTCCATCTCCACC |
| BS-36bpTR | AAACATTTAATAGACCTTAACTCCATCTCCACC-TexasRed |
| HS-20bpTR | GGCTTACCATTACCTCCACC-TexasRed |
| BS-31bp | AAACATTTAATAAACGCTTAACTCCATCTCCACC |
| BS-27bp | AACATTTAATAGACCTTAACTCCATCTCCACC |
| BS-21bp | GTATTTAATAAACGCTTAACTCCATCTCCACC |

The sequences of the respective forward oligonucleotides are shown. These oligonucleotides were hybridised with their complementary oligonucleotides and used as double-stranded substrates in EMSA experiments. BS-31bp contains one complete ORF80 binding site and was used for most experiments. The fluorescent dye Texas Red is attached at the 5′-end of BS-36bpTR. The asterisks above the DNA substrates indicate the region protected by ORF80 as revealed by DNase I footprinting. The bases in bold are conserved in both binding sites and are palindromic (see also Fig. 3C). HS-20bpTR contains only one half-site of the ORF80 binding site.

DNase I footprinting

Aliquots of 30 pmol footprint forward primer (5′-CGGCAGAGGCGCAAGG) were labelled with 6 pmol [γ-³²P]ATP (1.2 MBq; Hartmann, Germany) by incubation for 15 min with T4 polynucleotide kinase (Fermentas) at 37°C. The labelled primer was used directly in PCR with unlabelled reverse primer (5′-TGCGACACTTCAAGTCCGAC). The annealing temperature was 60°C and the template for amplification was pUC-pRN1. The resulting PCR fragment of 470 bp was pure as judged by denaturing gel electrophoresis and was purified by adsorption to silica (QiaQuick; Qiagen).
Aliquots of 1–5 pmol ORF80, 100 ng poly[d(I-C)·d(I-C)] and the labelled PCR fragment (∼5000 Bq, 125 fmol) were combined in a 4 µl binding reaction in 50 mM Tris–HCl, pH 7.0, and incubated for 5 min at 37 °C. Then 3 µl of 0.05 or 0.1 µM DNase I diluted in 20 mM Tris–HCl, 5 mM CaCl₂, pH 7.5, were added to the binding reaction. After 2 min at room temperature, the reaction was stopped by adding 12 µl of stop solution (80% formamide, 0.1% bromphenol blue, 0.1% xylene cyanol, 10 mM EDTA). A sample of 5 µl of the solution was incubated for 5 min at 80°C and then loaded on an 8% polyacrylamide sequencing gel. The dried gel was exposed to storage screens and analysed on a PhosphorImager (Molecular Dynamics).

Fluorescence titrations

All fluorescence titrations were carried out on a LS-50B fluorospectrometer from Perkin Elmer equipped with a thermostat and a stirred jacket. The analytical titrations were done in 1 ml cuvettes with 5 nM BS-36bpTR (Table 1) in standard buffer (25 mM sodium phosphate, 100 mM NaCl, 0.01% Tween-20, pH 7.0). The fluorescence of free DNA was recorded at 613 nm with a slit width of 20 nm (excitation at 585 nm with a slit width of 15 nm). ORF80 was added to a final concentration of 1 µM. Then, 300 µl of the solution in the cuvette was serially replaced (typically 12 times) by 300 µl of a 5 nM DNA solution in standard buffer. For the competition experiments 5 nM BS-36bp was allowed to bind to 250 nM ORF80. Then increasing amounts of unlabelled BS-36bp and of calf thymus DNA were added. For the non-specific competitor DNA we assumed that the number of binding sites equals the number of base pairs, implying overlapping binding sites, which is allowed as the number of binding sites exceeds by far the protein concentration.

For the stoichiometric titrations we used 0.5 µM BS-36bpTR and 0.5 µM HS-20bpTR (Table 1). The titrations were carried out in standard buffer and a 100 µl cuvette. For these measurements the same excitation and emission wavelengths were used, but the slit widths were narrowed to 5 and 6.5 nm, respectively.

**Dimethylsuberimidate (DMS) crosslinks**

Aliquots of 300 pmol ORF80 were incubated with and without 300 pmol double-stranded binding site DNA (BS-31bp) in the presence of 10 mM DMS in 0.2 M triethanolamine, pH 8.0, for 2 h at room temperature in a reaction volume of 30 µl. Aliquots of the reactions were separated on a 13% polyacrylamide peptide gel according to Schagger and von Jagow (9).

**RESULTS**

**Expression and purification**

ORF80 was overexpressed in *E.coli* using the T7 expression system. Purification of ORF80 proved to be difficult due to the
low expression of soluble ORF80 and its high tendency to aggregate. We finally worked out a two column strategy, which yielded essentially pure ORF80 (see Fig. 7, lane 1). From the retention of purified ORF80 in SDS–PAGE gels an apparent molecular mass of 11 kDa was determined, which agrees with the theoretical molecular mass of the full-length protein of 9.6 kDa. The N-terminus of the purified protein was determined to be GSHMSD, indicating that the initiator methionine is removed by the E.coli host cell during expression.

When purified ORF80 is heated for only 1 min at 95°C in the presence of 0.5% SDS and 0.25% β-mercaptoethanol prior to loading onto a denaturing peptide gel, a second protein band with an apparent molecular mass of 23 kDa is observed (data not shown). We ascribe this protein band to incompletely denatured dimers of ORF80. Unfortunately we could not analyse the quaternary structure of ORF80 by gel filtration as the protein aggregated on the column, even under slightly denaturing conditions (1 M urea). Additional evidence that ORF80 is dimeric in solution comes from our DMS and glutaraldehyde crosslinks (see below).

**ORF80 is a DNA-binding protein**

First, the DNA-binding activity of recombinant ORF80 was assayed with various restriction fragments of pUC18–pRN1. Figure 2 shows a restriction with Clal/SphI yielding a 472 bp fragment whose mobility decreases in the presence of ORF80. More importantly, when ORF80 was probed with a TaqI restriction digest that yields numerous restriction fragments only one fragment of ~610 bp was specifically shifted. This fragment and the 472 bp Clal/SphI fragment originate from the same region of pRN1 around the gene orf80 (Fig. 1B). The decrease in mobility after complexation with ORF80 is very large, suggesting that a large protein–DNA complex is formed. Indeed, binding of a single molecule of a 9.6 kDa protein to a DNA fragment of several hundred base pairs length is unlikely to result in a significant decrease in mobility of the complexed DNA compared with the unbound DNA.
To further narrow down the exact binding site of ORF80 we performed a DNase I footprinting assay using a probe encompassing this region. The footprint shown in Figure 3A clearly demonstrates that ORF80 binds to two sites within this probe. The two binding sites are separated by about 60 bases. At each binding site about 27 bases are protected against DNase I digestion. In addition, several hypersensitive DNase I sites are generated within each binding site. Figure 3B shows the precise location of the two binding sites relative to the initiator methionine of orf80. The two binding sites have a very similar structure. Each contains two TTAA motifs whose centres are separated by 11 bp, i.e. about one helix turn. Both binding sites have a pronounced palindromic structure. The TTAA motifs are themselves palindromic and two of the TTAA boxes together with four other nucleotides are arranged in a highly symmetrical manner (see boxed bases in Fig. 3C).

To study the DNA-binding properties of ORF80 in more detail we investigated the binding of ORF80 to short double- and single-stranded oligonucleotides whose sequences were derived from the binding site that is more distal to the presumed transcription start site. In binding reactions containing 500 nM ORF80 and a 31 bp double-stranded oligonucleotide that encompasses the complete binding site we observed a large protein–DNA aggregate that was not able to enter the gel (Fig. 4A, lane 2). When large amounts of competitor DNA, either calf thymus DNA or poly[d(I-C)]-d(I-C)], were added to the binding reaction, a defined protein–DNA complex was formed that was capable of entering the gel (lanes 3 and 7). In the presence of unlabelled binding site DNA this protein–DNA complex was no longer detectable (lane 4). These EMSAs clearly show that ORF80 binds specifically to short double-stranded DNA probes derived from the region protected against DNase I digestion.

In binding experiments with a 31 nt single-stranded probe derived from the binding site we observed the formation of a large complex that could not enter the gel when 500 nM ORF80 was present (Fig. 4A, lane 9). The addition of single-stranded competitor DNA (1.5 μM bases of M13 DNA) to this binding reaction completely abolished complex formation with the probe (lane 10), indicating that ORF80 does not bind to this single-stranded form of the binding site in a specific way.

In the absence of competitor DNA we always observed formation of a large protein–DNA complex, which was unable to enter the gel. This observation is certainly related to the high tendency of ORF80 to form aggregates, a property that was a severe obstacle in the purification scheme. Apparently, the presence of competitor DNA reduces the concentration of free ORF80 due to non-specific affinity of ORF80 for the competitor.

In order to observe defined protein–DNA complexes in the gel shift assays we therefore routinely included 5 μM base pairs of poly[d(I-C)]-d(I-C)] as competitor. In the presence of this amount of competitor DNA half-maximal binding to the double-stranded binding site was achieved at ~700 nM ORF80 (Fig. 5A). Quantitative analysis of the binding isotherm suggests that the binding is cooperative, with a Hill constant of ~2 (Fig. 5C).

In further gel shift experiments we investigated the minimal size of the binding site that allows specific binding in the presence of competitor DNA. In the DNA probes of 36, 31, 27 and 21 bp length the binding site is stepwise shortened (Table 1). A distinct protein–DNA complex was observed for the 36 and 31 bp probes, whereas with the smaller probes only a small amount of aggregate was formed (Fig. 4B). This observation is in line with the footprinting data, as the 31 bp substrate just covers the protected region (Table 1), indicating that the whole footprinting area is required for specific binding. Deletion of the border regions of the footprinting area apparently leads to a loss of contacts that are necessary for formation of a stable specific complex.

The footprinting data have revealed the existence of two ORF80 binding sites that are separated by ~60 bp. In order to address the question of whether there is cooperativity between the adjacent binding sites, we prepared a 149 bp PCR fragment carrying both sites and analysed ORF80 binding to this probe by gel shift titration experiments. As shown in Figure 5B,
several complexes were formed in a cooperative manner. At high concentrations of ORF80 the probe was saturated and a distinct protein–DNA complex of rather low mobility was formed. Intermediary complexes were also visible; however, these complexes were only weakly populated. Analysis of these titrations suggests that the cooperativity of this binding reaction has a Hill coefficient of $\sim 4$ (Fig. 5C), which gives the lower limit of the number of ORF80 molecules that bind cooperatively to the two sites. This result is in line with the large decrease in mobility of the restriction fragments upon ORF80 binding (see Fig. 2).

Fluorescence titrations

To study more precisely the binding affinity and stoichiometry of the DNA–ORF80 protein complex, we developed a fluorescence-based DNA-binding assay for ORF80. We discovered that BS-36bp labelled with the fluorescent dye Texas Red at the 3'-end of the forward strand is a suitable probe for fluorescence titration. As can be seen in Figure 6A, the fluorescence of Texas Red is quenched upon ORF80 binding to about one-third of the initial value. At concentrations >500 nM a slight increase in fluorescence intensity was observed, which is ascribed to the formation of aggregates and/or complexes of higher molecular order. The lower part of the titration, however, could be very well fitted to the Hill equation and yielded a Hill coefficient of $n_{\text{Hill}} = 1.9$ with half-maximal binding at 136 nM. The observed fluorescence intensity change is not caused by simple binding of ORF80 to the dye, as revealed by competition experiments with BS-36bp (specific DNA) and with calf thymus DNA (non-specific DNA). As expected, the specific DNA was a good competitor, whereas high amounts of the non-specific DNA are needed to displace ORF80 from the fluorescently labelled specific DNA probe (Fig. 6B). The competition experiments also clearly demonstrate that the type of inhibition is different between the specific DNA and the non-specific competitor. The slope of the curve is much steeper for the specific DNA, indicating that the specific competitor binds ORF80 in a cooperative manner. The rapid increase in fluorescence intensity from 20 to 40 nM competitor indicates a rapid decrease in free ORF80 in this concentration range caused by cooperative binding of ORF80 to the specific competitor.

In order to determine the stoichiometry of the protein–DNA complexes we performed fluorescence titrations with 0.5 µM DNA. Under these conditions we expected a nearly linear decrease in fluorescence intensity until the DNA probe was completely saturated with protein. With the half-site binding DNA (HS-20bpTR) stoichiometric equivalence was reached at a protein/DNA ratio of about six. Twice the amount, i.e. about 12 monomers of ORF80, are required to saturate the DNA with one complete binding site (BS-31bpTR). We have no indication whether the protein is fully active. However, we obtained similar titration curves with different protein preparations. The experiment therefore further supports the hypothesis that a rather large complex consisting of several ORF80 monomers is formed on the binding site DNA (see also Discussion).

Oligomer structure of ORF80 and ORF80–DNA complexes

To study the oligomer structure of free and bound ORF80 we performed crosslinking experiments with the bifunctional amine-reactive reagent DMS. When ORF80 was allowed to react with DMS in the absence of binding site DNA, we observed an additional protein band with an apparent molecular mass of 24 kDa in denaturing peptide gels (Fig. 7, lane 2). This protein band is ascribed to a crosslink between two ORF80 molecules. In crosslinking reactions in the presence of DNA the dimeric protein band was stronger and in addition protein bands corresponding to crosslinking of three and four ORF80 polypeptide chains are visible (lane 4). The same pattern is
seen when crosslinking was done with glutaraldehyde (data not shown). These experiments indicate that ORF80 is a dimer in solution and assembles into a larger complex when bound to DNA.

Is ORF80 a basic leucine zipper?

Analysis of the primary structure of ORF80 revealed several important properties (Fig. 8). The 30 N-terminal amino acids of ORF80 have a high coiled coil score (10) and are predicted to be helical. In addition, there are five leucines completely conserved among the ORF80 homologues from pRN1, pRN2, pDL10, pSSVX, pING1 and SSO8948. Four of them are in the correct register to form a leucine zipper (positions a and d in a helical wheel presentation). The ORF80 homologue from pNOB8 has an isoleucine in one of the five conserved leucine positions and the homologue from pHEN7 has a methionine in one position. In contrast, two of the genomic ORF80 homologues (SSO8758 and SSO5209) have none or two of these leucines conserved and might therefore be unable to form a leucine zipper. These data strongly suggest that the plasmid subsets of ORF80 homologues as well as SSO8949 can multimerise via the N-terminal domains to form a leucine zipper. For ORF80 the existence of dimers is demonstrated by SDS–PAGE experiments and chemical crosslinking. CD measurements allow an estimation of the secondary structure composition of a protein. The negative ellipticity in the wavelength region between 210 and 250 nm reveals that ORF80 has a significant fraction of helical regions (Fig. 9). Deconvolution of the CD spectrum with a neural network (11) yields an α-helical content of 36% and a β-strand content of 17.1%. These values agree quite well with the GOR IV predictions (43 and 13%, respectively) and thus further support the hypothesis that the N-terminus could be helical, which is a requirement for formation of a leucine zipper.

C-terminal to the putative leucine zipper domain there is a less conserved region of about 10 amino acids, which is followed by a highly conserved and positively charged region extending up to the C-terminus of the protein (Fig. 8). The charge and high conservation of this domain suggest that it functions as a DNA-binding region. As compared to the classic basic leucine zipper proteins, such as the transcription factor GCN4, where the basic region is N-terminal to the leucine zipper, these domains are arranged in reverse order in ORF80. Therefore, ORF80 could constitute a new subclass of leucine zipper DNA-binding proteins.

Figure 6. Fluorescence titrations with Texas Red-labelled binding site DNA. (A) Aliquots of 5 nM Texas Red-labelled BS-36bp were titrated with increasing amounts of ORF80. The solid line is a fit according to the Hill equation \( n_{\text{Hill}} = 1.93, k_{\text{Hill}} = 138 \text{ nM} \). (B) Competition experiments with 250 nM ORF80 bound to 5 nM DNA. Increasing amounts of competitor DNA were added. Filled circles, unlabelled specific DNA (BS-36bp); open circles, calf thymus DNA. (C) Stoichiometric titration with 0.5 μM Texas Red-labelled BS-36bp (filled circles) and with HS-20bp (open circles). Stoichiometric equivalence was reached at protein/DNA ratios of about 12 and 6, respectively.

Figure 7. Crosslinking of ORF80 with DMS. ORF80 was incubated with and without DMS and analysed on a denaturing peptide gel. Lane 1, no DMS; lane 2, 10 mM DMS added. Chemical crosslinking was repeated in the presence of binding site DNA (BS-31bp). Lane 3, no DMS; lane 4, 10 mM DMS.
All ORF80 homologues have a highly conserved and highly basic C-terminal domain, which is presumed to be involved in DNA binding. Given the high conservation of this putative DNA-binding domain we searched the nucleotide sequence of the respective plasmids for ORF80 binding sites. Upstream of all orf80 genes we find two potential ORF80 binding sites that are all separated by the same distance (Fig. 10). However, the distances between the binding sites and the start codons differ. On plasmids pRN1, pRN2, pHEN7 and pSSVx and on the genomic copy SSO5209 binding of ORF80 to the less upstream binding site could occlude the BRE and TATA box (see also Fig. 3). Both elements are required for transcription (12,13). It therefore seems plausible that ORF80 down-regulates its own synthesis on these plasmids. For plasmids pDL10 and pING1, as well as for the genomic copies SSO8948 and SSO8758, the distance is larger and it is less likely that binding of ORF80 to these sites directly influences transcription of orf80. The most remarkable feature of the ORF80 binding sites is the strong conservation of the spacing of the two binding elements, indicating a close functional and/or structural relationship between the two sites.

DISCUSSION

The cryptic high copy number plasmid pRN1 from S.islandicus harbours three genes that are conserved among the plasmid family pRN. One of these genes, orf80 from pRN1, has no homology to other known proteins. It is, however, related to genes of the pRN family as well as to other Sulfolobus spp. plasmids. In fact, the orf80 gene is found in all Sulfolobus plasmids characterised so far, suggesting that orf80 codes for an essential plasmid function required for replication and maintenance of the plasmids in the host cell. It is noteworthy that ORF80 homologues were found neither outside the genus Sulfolobus (except pDL10) nor in Sulfolobus virus sequences. orf80 genes are found on the conjugative plasmids pING and pNOB8, suggesting that orf80 is not only restricted to the pRN plasmid family but is also required by conjugative plasmids from Sulfolobus strains. The orf80 gene has not been annotated in the virus–plasmid hybrid pSSVx, but the high similarity of the gene product and the conserved upstream region with the putative ORF80 binding sites strongly suggest that the orf80 homologue is also functional in this plasmid. Searching the complete genome sequence of S.solfataricus P2 we discovered three further homologues of orf80. The gene SSO8948 has high similarity to the plasmid orf80 genes, whereas the remaining two genes, SSO8758 and SSO5209, are more distantly related. The latter two genes code for proteins that do not possess the highly conserved leucine residues and the upstream region of SSO5209 does not contain the typical ORF80 binding site motif TTAA(N)7TTAA. It is unknown whether the three genes are functional and what function they may carry out. Given the high plasticity of the Sulfolobus
This point we cannot provide firm experimental data for this observed smeared shifted bands in the EMSA experiments. At have a defined stoichiometry. Otherwise we would have protein–DNA complex is nonetheless specific and seems to ORF80 molecules are only bound to the complex via protein–aggregation tendency of ORF80 it is not unlikely that some complexes probably not all ORF80 molecules can make direct with two TTAA(N)7TTAA motifs. In these protein–DNA motif and assemble into a protein–DNA complex with eight dimers could bind to short DNA sequences containing a TTAA consists of a sequence-specific double-stranded DNA-binding protein. Of a large number of TaqI restriction fragments derived from pRN1 only one was bound by ORF80 in EMSA. This fragment encompasses the complete gene orf80. Our footprinting analysis narrowed the binding site down to two sites of similar structure upstream of the transcription start site. These sites are highly conserved upstream of all plasmid orf80 genes. Fluorescently labelled short DNA comprising the complete protected region of the binding site were of an unusual structure and symmetry. In addition to two TTAA motifs, two other conserved nucleotides are arranged in a palindromic way. The palindromic structure suggests that an ORF80 oligomer of 2-fold symmetry is the active DNA-binding species. A search of the TFSITE database did not reveal a transcription factor that binds to a bipartite TTAA motif. The EMSA experiments and the site titrations suggest that a function of the observed cooperativity in protein binding that leads to specific structural role for these complexes. Possibly the function of the conserved second binding site more a palindromic way. The palindromic structure suggests that an ORF80 oligomer of 2-fold symmetry is the active DNA-binding species. A search of the TFSITE database did not reveal a transcription factor that binds to a bipartite TTAA motif. The EMSA experiments and the site titrations suggest that a large but specific protein–DNA complex is formed when ORF80 binds to its target DNA. It is possible that two ORF80 dimers could bind to short DNA sequences containing a TTAA motif and assemble into a protein–DNA complex with eight ORF80 subunits and a TTAA(N)7TTAA motif. Consequently, 16 or more subunits would then form a complete binding site with two TTAA(N),TTAA motifs. In these protein–DNA complexes probably not all ORF80 molecules can make direct DNA contacts to the rather short DNAs. Given the high aggregation tendency of ORF80 it is unlikely that some ORF80 molecules are only bound to the complex via protein–protein interaction. It should be stressed, however, that the protein–DNA complex is not sequence specific and seems to have a defined stoichiometry. Otherwise we would have observed smeared shifted bands in the EMSA experiments. At this point we cannot provide firm experimental data for this binding model but our site titrations suggest that about six subunits are required to saturate a single TTAA motif and that about 12 subunits can bind to a TTAA(N)2,TTTAA motif. These values are based on the assumption that the protein is fully active. The protein for these experiments was prepared according to the alternative purification scheme during which the protein was purified in the presence of 6 M urea. As this amount of urea does not unfold the protein substantially (9.7 M urea was needed to reduce the CD signal to 50%; Fig. 9) and as the CD spectrum of ORF80 was reasonable it can be assumed that most of the protein was folded and active in these assays.

What might be the physiological role of ORF80? ORF80 may have a structural and/or regulatory function. One of the ORF80-binding sites is close to the initiator methionine and indeed encompasses the archaean TATA box. It can therefore be speculated that ORF80 represses its own transcription in an autoregulatory manner. This bacterial-like transcriptional repression has recently been reported for the Lrs14 protein from S.solfataricus (15). However, on plasmids pDL10 and pING the binding sites are more distal to the initiator methionine, and the putative archaean TATA box of these two genes might not be blocked by binding of ORF80 to its binding site.

The function of the conserved second binding site more upstream of orf80 is even more unclear. In view of the strict conservation of the distance between the two sites and in view of the observed cooperativity in protein binding that leads to formation of a large protein–DNA complex, we assume a specific structural role for these complexes. Possibly the ORF80 complexes impart a distinct structure that is necessary for recruitment of, for example, the replication initiation machinery. Several authors have suggested that the region upstream of orf80 contains the double-stranded origin of replication (2,5). Experimental evidence supporting this hypothesis is, however, still lacking.

In preliminary experiments we asked whether partially purified recombinant ORF904, the presumed replication initiator protein, is able to interact with the ORF80–DNA complex. In EMSA, however, we were unable to demonstrate a supershift of the ORF80–DNA complex in the presence of ORF904 (G.Lipps, unpublished results).

Sequence analysis of the orf80 coding region suggests that dimethylation is mediated by a leucine zipper motif. Remarkably, as compared to all other known basic leucine zipper proteins, the basic domain and the leucine zipper are swapped in ORF80. Our present work suggests that both orientations

Figure 10. Alignment of the ORF80 putative binding sites. The promoter sequences were aligned with PILEUP (GCG program package) and the consensus calculated with PRETTY. The highly conserved TTAA motifs are boxed. The centres of the two TTAAs motifs within one binding site are separated by 11 bp (Fig. 9) and as the CD spectrum of ORF80 was reasonable it can be assumed that most of the protein was folded and active in these assays.
allow specific recognition of DNA. Basic leucine zippers have mostly been found in eukaryal transcription factors. To our knowledge the transcriptional activator GvpE of Halobacterium salinarum has up to now been the only archaeal protein with a basic leucine zipper motif (16). In contrast to GvpE, which is a 21 kDa protein, ORF80 is of smaller size and seems to have no additional binding site for ligands that could regulate its function. ORF80 may therefore represent a minimal model protein for this new subclass of basic leucine zipper proteins.

ACKNOWLEDGEMENTS

We thank Prof. W. F. Doolittle (Dalhousie University, Halifax) for plasmid pUC18-pRN1 and Prof. Garrett (Copenhagen University) for the sequence of the plasmid pHEN7 prior to publication. We are also grateful to our colleague T. Hey for critical reading of the manuscript.

REFERENCES