Article No. mb981795 *J. Mol. Biol.* (1998) **279**, 761–771

JMB

The Transcriptional Activator GvpE for the Halobacterial Gas Vesicle Genes Resembles a Basic Region Leucine-zipper Regulatory Protein

Kerstin Krüger[1], Thomas Hermann[2], Vanessa Armbruster[1] and Felicitas Pfeifer[1]*

[1]*Institut für Mikrobiologie und Genetik, Schnittspahnstr. 10 Technische Universität Darmstadt, D-64287 Darmstadt, Germany*

[2]*Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany*

**Corresponding author*

The GvpE protein involved in the regulation of gas vesicles synthesis in halophilic archaea has been identified as the transcriptional activator for the promoter located upstream of the *gvpA* gene encoding the major gas vesicle structural protein GvpA. A closer inspection of the GvpE protein sequence revealed that GvpE resembles basic leucine-zipper proteins typically involved in the gene regulation of eukarya. A molecular modelling study of the C-terminal part implied a cluster of basic amino acid residues constituting the DNA-binding site (DNAB) followed by an amphiphilic helix, suitable for the formation of a leucine-zipper structure within a GvpE dimer. The model of a GvpE dimer docked onto DNA indicated that the side-chains of the basic residues could perfectly interact with the negatively charged phosphate groups of the DNA backbone. Substitution of three basic amino acid residues of this putative DNAB by alanine and/or glutamate generated mutated GvpE proteins. None of these was able to activate the c-*gvpA* promoter *in vivo*, indicating that these basic residues are required for GvpE activity. This identification of an archaeal gene regulator displaying similarity to eukaryal regulatory proteins implies that the basic transcription machinery of eukarya and archaea are closely related, and that the regulatory proteins have evolved according to common principles.

Keywords: gas vesicles; archaea; transcription regulation; bZIP protein

Introduction

Archaea constitute the third domain of life according to the Woesian tree (Woese *et al*., 1990). These organisms share a common origin with eukarya, which is most obvious when enzymes involved in gene expression are compared. Archaeal RNA polymerases consist of eight to 13 protein subunits, which are more similar to the respective subunits of eukaryal RNA polymerases than to the bacterial RNA polymerase (Langer *et al*., 1995). In addition, archaeal promoters contain a TATA-box sequence located approximately 27 nucleotides upstream of the transcription initiation site (Hain *et al*., 1992; Langer *et al*., 1995). Like in eukarya, transcription initiation involves the action of TATA-box binding proteins (TBP) and transcription factors (TFII), but the mechanism appears to be somewhat simpler, since the archaeal system requires only the homologs of the eukaryal TBP and TFII B for accurate initiation, and the TFII E and TFII H required for eukaryal transcription initiation are not present (Thomm, 1996; Reeve *et al*., 1997; Qureshi *et al*., 1997). While the basic archaeal transcription machinery has been investigated in some detail, insight into the regulation of gene expression in archaea is still scarce. In the case of halophilic archaea, the molecular tools for the investigation of gene functions are well developed. Several vector plasmids, including expression vectors, and a transformation system

Present address: T. Hermann, Institut de Biologie Moléculaire et Cellulaire (IBMC), UPR9002 du CNRS, 15 rue René Descartes, F-67084 Strasbourg Cedex, France.

Abbreviations used: TBP, TATA-box binding protein; TF, transcription factor; bZIP, basic leucine-zipper protein; DNAB, DNA binding site.

0022–2836/98/240761–11 $25.00/0

are available and have been used for the study of promoters and gene expression in halobacteria (Gropp *et al*., 1995; Pfeifer *et al*., 1994; Palmer & Daniels, 1995; Danner & Soppa, 1996; Röder & Pfeifer, 1996).

Gas vesicle formation in halobacteria is a rewarding system to study archaeal gene regulation. These gas-filled, proteinaceous structures synthesized by *Halobacterium salinarum* and *Haloferax mediterranei* allow the organism to maintain buoyancy. In the case of *Hf. mediterranei* they appear solely in the stationary growth phase and under conditions of high salinity (Englert *et al*., 1990; Röder & Pfeifer, 1996). In contrast, *Hb. salinarum* PHH1 wild-type produces gas vesicles constitutively during all stages of growth, where as a mutant strain, *Hb. salinarum* PHH4, forms gas vesicles exclusively in the stationary growth phase (Horne & Pfeifer, 1989; Horne *et al*., 1991). In each case, gas vesicle formation requires 14 different *gvp* (gas vesicle protein) genes, which are arranged as two clusters in the so-called vac region, namely *gvpACNO*, and upstream of *gvpA*, oriented in opposite direction, *gvpDEFGHIJKLM* (Englert *et al*., 1992a). The wild-type strain *Hb. salinarum* PHH1 contains the 14 *gvp* genes responsible for the constitutive gas vesicle formation on plasmid pHH1. The p-*gvpA* gene encoding the major gas vesicle structural protein is transcribed from the pA promoter (Offner & Pfeifer, 1995; Offner *et al*., 1996), whereas the chromosomally located *gvpA* gene in *Hb. salinarum* PHH4 (c-*gvpA*), and in *Hf. mediterranei* (mc-*gvpA*) are transcribed in stationary growth phase only (Krüger & Pfeifer, 1996; Röder & Pfeifer, 1996).

In contrast to the pA promoter, the mcA and cA promoters are inactive during exponential growth of *Hf. mediterranei* and *Hb. salinarum* PHH4, respectively, and are activated during gas vesicle formation (Horne & Pfeifer, 1989; Englert *et al*., 1990). Interestingly, a *Hf. volcanii* transformant containing the mc-*gvpA* gene by itself synthesizes mc-*gvpA* mRNA at a very low level compared to the normal expression, indicating that the mcA promoter needs activation (Englert *et al*., 1992b; Röder & Pfeifer, 1996). Further investigations demonstrated that the product of mc-*gvpE*, mc-GvpE, could establish mcA activation (Röder & Pfeifer, 1996).

Here, we describe an investigation of the cA promoter and characterize the transcription activator cGvpE in more detail. Transformation experiments and subsequent RNA analyses demonstrated that the cA promoter absolutely requires activation by cGvpE. Analysis of the amino acid sequence of GvpE and a subsequent molecular homology modelling approach suggested that the protein resembles eukaryal basic leucine-zipper (bZIP) proteins, which act as transcription regulators such as cFos, cJun and GCN4 (Ellenberger, 1994). A high level of structural similarity to the yeast GCN4 bZIP protein (O'Shea *et al*., 1991; Ellenberger *et al*., 1992) was found in the leucine-zipper part of GvpE, which is highly conserved among the various GvpE proteins. In order to assess the prediction that GvpE function depends on a basic leucine-zipper structure, the putative cGvpE-DNA binding motif was altered by mutagenesis at residues determined from the molecular model, and the effects of these various GvpE mutations on c-*gvpA* expression were studied *in vivo*.

Results

Analyzes of the pA and cA promotes in transformants

The activity of the p-*gvpA* (pA) and c-*gvpA* (cA) promoter elements was investigated using the halobacterial shuttle vector pJAS33. This vector allows the validation of promoter-containing fragments *via* the expression of the dihydrofolate reductase (*dhfr*) reading frame (Zusman *et al*., 1989). The ferredoxin promoter (Pfeifer *et al*., 1993), inserted 5′ to *dhfr*, can be replaced by other DNA fragments *via Pst*I and *Nco*I sites.

Fragments containing the pA or cA promoter were inserted upstream of the *dhfr* reading frame, and *Hf. volcanii* was transformed with the resulting constructs. The presence of the *dhfr* transcript in each of these transformants was estimated by Northern analysis, which showed that the original ferredoxin promoter as well as the pA promoter led to the synthesis of a 150 nt RNA species covering the *dhfr* reading frame and additional sequences downstream (Figure 1(a)). The cA promoter fragment in this construct, however, led to the production of a 2.7 kb RNA (Figure 1(a)). The exact start site of each of these transcripts was determined by primer extension analysis (Figure 1(b)). In the case of the *fdx* promoter in pJAS33, the signal observed located the transcript start 25 nt downstream of the Box A element (Figure 2). This distance of the start site was the same as determined for the intact *fdx* gene (Pfeifer *et al*., 1993). In the case of the pA promoter, the transcript started 24 nt downstream of the Box A element (Figure 2, pA). Again, the start site appeared at the same distance as determined for the intact p-*gvpA* gene, although the sequence around the transcription start site was altered in the pJAS construction (see Figure 2; Horne & Pfeifer, 1989). These two primer extension signals showed the same relative strength as the signals obtained by Northern blot analysis (Figure 1(a)), indicating that the promoter activity of the ferredoxin promoter exceeds that of the pA promoter fragment. In contrast, the RNA sample of the transformant containing the cA promoter construct did not give rise to a primer extension signal, indicating the lack of a defined start site of the 2.7 kb RNA in the investigated region (Figure 1(b)), and data not shown). Thus, the cA promoter is not active in *Hf. volcanii*, and the observed 2.7 kb tran-

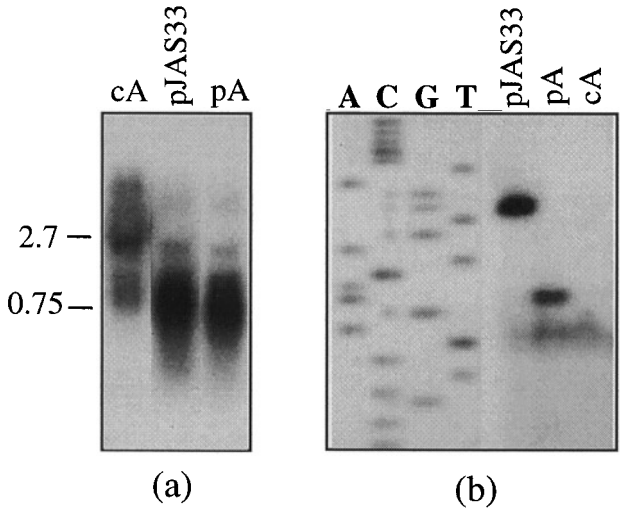

Figure 1. (a) Northern analyses of the *Hf. volcanii* transformants carrying the constructs cA×pJAS33 (cA), pJAS33 and pA×pJAS33 (pA) using the *dhfr* gene as probe: 3 μg of total RNA was isolated from *Hf. volcanii* transformants and separated on formaldehyde/agarose gels. Numbers on the left indicate the size of the hybridizing mRNA (in kilobases). (b) Primer extension analyses of the same *Hf. volcanii* transformants. The analysis was done with 5 μg of total RNA and 0.5 pmol of labelled oligonucleotide. The marker sequence (ACGT) was obtained with the T3 primer and pBluescript SK^+ DNA.

script started at a fortuitous initiation site in the vector.

Transformation experiments to detect the *gvp* gene required for the cA promoter activation

Since the cA promoter element was not sufficient to drive the expression of the *dhfr* reading frame, transformation experiments were conducted to identify the element promoting c-*gvpA* transcription. *Hf. volcanii* was transformed with (i) the A1 construct containing the c-*gvpA* gene including its own promoter, and (ii) the A1 construct together with a plasmid carrying the c-*gvpE-M* genes of the c-vac region (E-M construct, see Figure 3(a)). The expression of c-*gvpA* in these transformants was tested by Northern analysis using a c-*gvpA* specific probe, and in Western analysis with an anti-gas vesicle serum detecting the GvpA monomer (Englert *et al.*, 1992b). The isolated RNA of the A1 transformant did not hybridize with the c-*gvpA* specific probe (Figure 3(b)), demonstrating that the cA promoter is not sufficient for transcript initiation. Consequently, lysates of two different A transformants (A1 and A2, see Materials and Methods) also did not react with the anti-gas vesicle serum (Figure 3(c), GvpA). The A1/E-M transformant, however, contained c-*gvpA* mRNA (Figure 3(b)) as well as the cGvpA protein, as shown by Western analysis (Figure 3(c), GvpA), suggesting that the cA promoter was activated *in trans* by at least one of the products of the c-*gvpE-M* genes. Further analyses led to the identification of cGvpE as the transcriptional activator, since c-*gvpA* expression was observed in a double transformant carrying the c-*gvpA* gene (A2 construct) and the c-*gvpE* reading frame in the expression vector pJAS35 under *fdx* promoter control (Eexp; Figure 3(c), GvpA). Also, the amount of cGvpE protein was analyzed in these transformants using an antiserum raised against isolated cGvpE synthesized in *Escherichia coli* (anti-cGvpE serum; Krüger & Pfeifer, 1996). Each transformant carrying the c-*gvpE* gene (A1/E-M, Eexp and A2/Eexp) contained the cGvpE protein (Figure 3(c), GvpE), whereas lysates of *Hf. volcanii* and the transformants containing the c-*gvpA* gene by itself (A1 and A2) did not react (Figure 3(c), GvpE).

Molecular modelling of the GvpE protein

Comparison of the full-length GvpE primary structure against the protein sequence database did not reveal similar proteins. However, inspection of the GvpE sequence revealed a leucine-rich region in the C-terminal part and a cluster of basic lysine and arginine residues around position 145 (Figure 4), both characteristic features of the basic leucine-zipper motif in eukaryal transcription factors (Hu *et al.*, 1990). Secondary structure prediction for GvpE suggested a long amphiphilic helix of more than 30 amino acid residues (AH6, Figure 4), starting at position 152, which is centered around a tandem of leucine residues (L165 and L172) at a distance of seven residues. The hydrophobic residues valine (cGvpE and mcGvpE) or isoleucine (pGvpE) are found at position 158, seven residues from the N terminus of the leucine tandem. Also, a conserved cysteine residue C179, occurs seven residues after the leucine tandem towards the C terminus (Figure 4). Cysteine replacement for leucine in a leucine-zipper is unusual; however, replacement by residues of even higher polarity such as threonine occurs in eukaryal bZIP proteins (Pabo & Sauer, 1992). All the hydrophobic residues in AH6 form a characteristic 4,3 repeat and cluster on one side of the putative amphiphilic helix with the four key residues V/I158, L165, L172 and C179 lined up along the center of the hydrophobic helix surface. The hydrophobic residues in AH6 are conserved within GvpE proteins of the three vac regions, along with two acidic residues at positions 159 and 178. The N terminus of the putative helix AH6 is flanked by a cluster of four basic residues (K142 to R147), which is predicted to have no defined secondary structure, while another α-helix of 19 residues (H5, Figure 4) starts immediately N-terminal to the basic cluster.

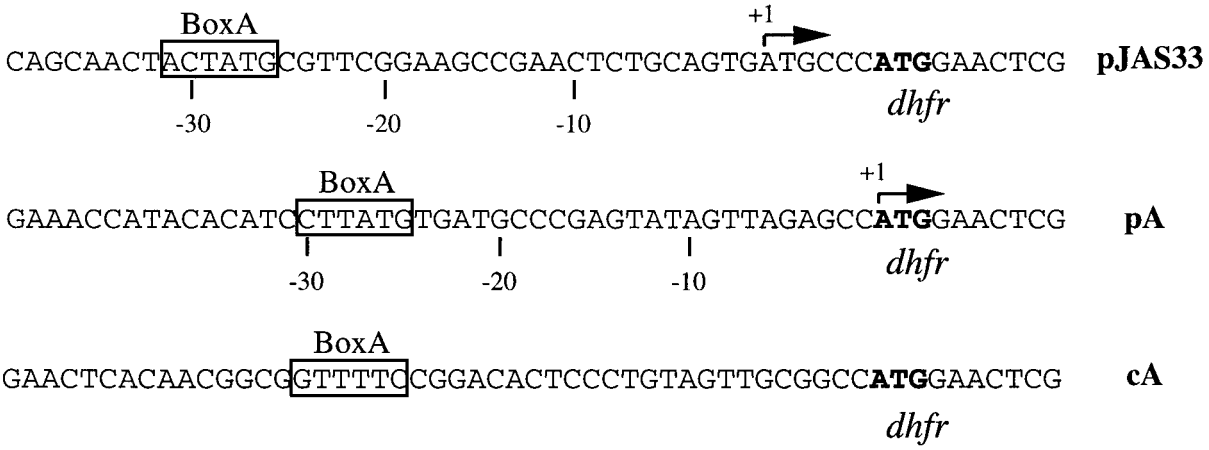

Figure 2. DNA sequences of the promoters inserted in front of the *dhfr* reading frame in pJAS33, pA×pJAS33 (pA) and cA×pJAS33 (cA). The promoter BoxA elements are marked by a box surrounding the sequence. The ATG start codon of the *dhfr* reading frame is printed in bold. +1 above an arrow indicates the transcriptional start site as determined by primer extension analysis.

Besides the two helices H5 and AH6, four other regions of GvpE were predicted to form α-helices designated AH1, H2, AH3 and AH4, where A indicates an amphiphilic helix structure (Figure 4). Comparison of sequence fragments of the six helices and the cluster of basic residues with the Protein Database revealed similar sequences for AH1, AH6 and the basic cluster. The putative helix AH1 shows a high level of similarity to an α-helix of the Ca^{2+}-binding motif within the class of annexins, Ca^{2+}-dependent phospholipid-binding proteins (Smith & Moss, 1994 and see Figure 4). Helix AH3 contains a region of eight consecutive hydrophobic residues corresponding to two helix

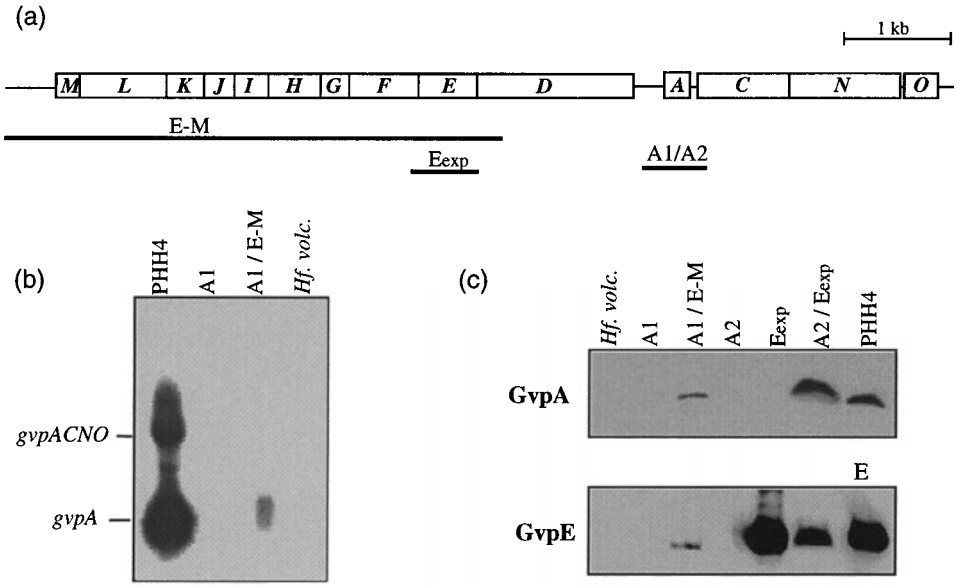

Figure 3. Constructs used for transformation experiments, and Northern and Western analyses. (a) Each *gvp* gene of the c-vac region is represented by a box labelled A and C through O. The subfragments in constructs are indicated as black bars underneath. The A1 and A2 constructs contain the same c-*gvpA* fragment, but ligated to different halobacterial vectors. (b) Northern analysis of the *Hf. volcanii* transformants A1 and A1/E-M, *Hb. salinarum* PHH4 and *Hf. volcanii* using a c-*gvpA*-specific probe. In each case, 5 μg of total RNA was separated on a formaldehyde/agarose gel. The *gvpA* and *gvpACNO* mRNAs of *Hb. salinarum* PHH4 (Krüger & Pfeifer, 1996) are labelled. (c) Western analyses of the *Hf. volcanii* transformants and the respective controls. The analysis was done with 20 μg of protein each, and anti-gas vesicle serum (labelled GvpA) or anti-cGvpE serum (GvpE) was applied. The lane designated PHH4 contains protein isolated from a stationary culture from *Hb. salinarum* PHH4, the lane designated E contains isolated cGvpE protein synthesized in *E. coli* (Krüger & Pfeifer, 1996).

Figure 4. Alignment and secondary structure prediction of cGvpE, pGvpE, mcGvpE and parts of other proteins sharing similar domains. Underneath the GvpE alignment a consensus sequence is given depicting conserved, acidic (a) or hydrophobic (h) amino acid residues. Regions within GvpE for which α-helical secondary structure is predicted are underlined and labelled H1 through AH6, with amphiphilic helices designated AH. The hydrophobic core residues (bold) within the leucine-zipper region (AH6) are marked by dots and the putative DNA binding motif (with the basic amino acid residues given in bold) by DNAB. Below the region of H5 to AH6, part of the eukaryal bZIP transcription factor GCN4 sequence is given. Above AH1 an alignment of a region from two different annexins (Anx) is shown that forms an α-helix participating in the Ca^{2+}-binding domain of these proteins (Lewit-Bentley *et al.*, 1992). The α-helix of chicken annexin V shares 67% identity and 92% similarity with the GvpE region, that of human annexin V 58% identity and 92% similarity based on the evolutionary distances according to Dayhoff (1978).

turns, which are probably buried from solvent access.

The location of an unstructured cluster of basic residues between helices H5 and AH6, and the distribution of hydrophobic residues within AH6 led to the speculation that GvpE might constitute a basic leucine-zipper protein. Consequently, the active form of GvpE is predicted to be a dimer held together by the hydrophobic interactions between parallel oriented amphiphilic AH6 helices. We further hypothesized that one of the four Leu-Leu interactions that normally occur at the core of the leucine-zipper dimer interface is replaced by two facing cysteine residues. The basic residues at positions 142 to 147 (KRKYVR) could be engaged in binding of the putative GvpE dimer to DNA. Similar to the yeast GCN4 bZIP transcription factor (König & Richmond, 1993), this basic region (designated DNAB) might become helical only after binding to DNA, as indicated by the lack of predicted secondary structure.

To explore further the likelihood of our hypothesis, we constructed a three-dimensional model of the putative GvpE bZIP region comprising residues H125 to N182 (Figure 5(a)) as described in Materials and Methods using the crystal structure of the yeast bZIP protein GCN4 (Ellenberger *et al.*, 1992; König & Richmond, 1993) as a template. The respective GvpE sequence could be easily accommodated to the bZIP dimer structure, with the side-chains neatly fitting the hydrophobic interface between the coiled helices AH6 (Figure 5(a)).

The GvpE bZIP model was docked to the major groove of a regular *B*-form DNA, guided by the coordinates of a GCN4/DNA complex (König & Richmond, 1993), to study possible interactions between the basic region (DNAB) and the nucleic acid (Figure 5(b)). The GvpE/DNA model revealed that the side-chains of basic residues K142, R143, K144 and R147 within DNAB occupy positions perfectly suitable for interaction with the negatively charged phosphate groups of the DNA backbone. A striking difference is observed between the regions of the DNA that interact with the DNAB structures of GCN4 as compared to the modelled GvpE. While in GCN4, the basic DNAB side-chains

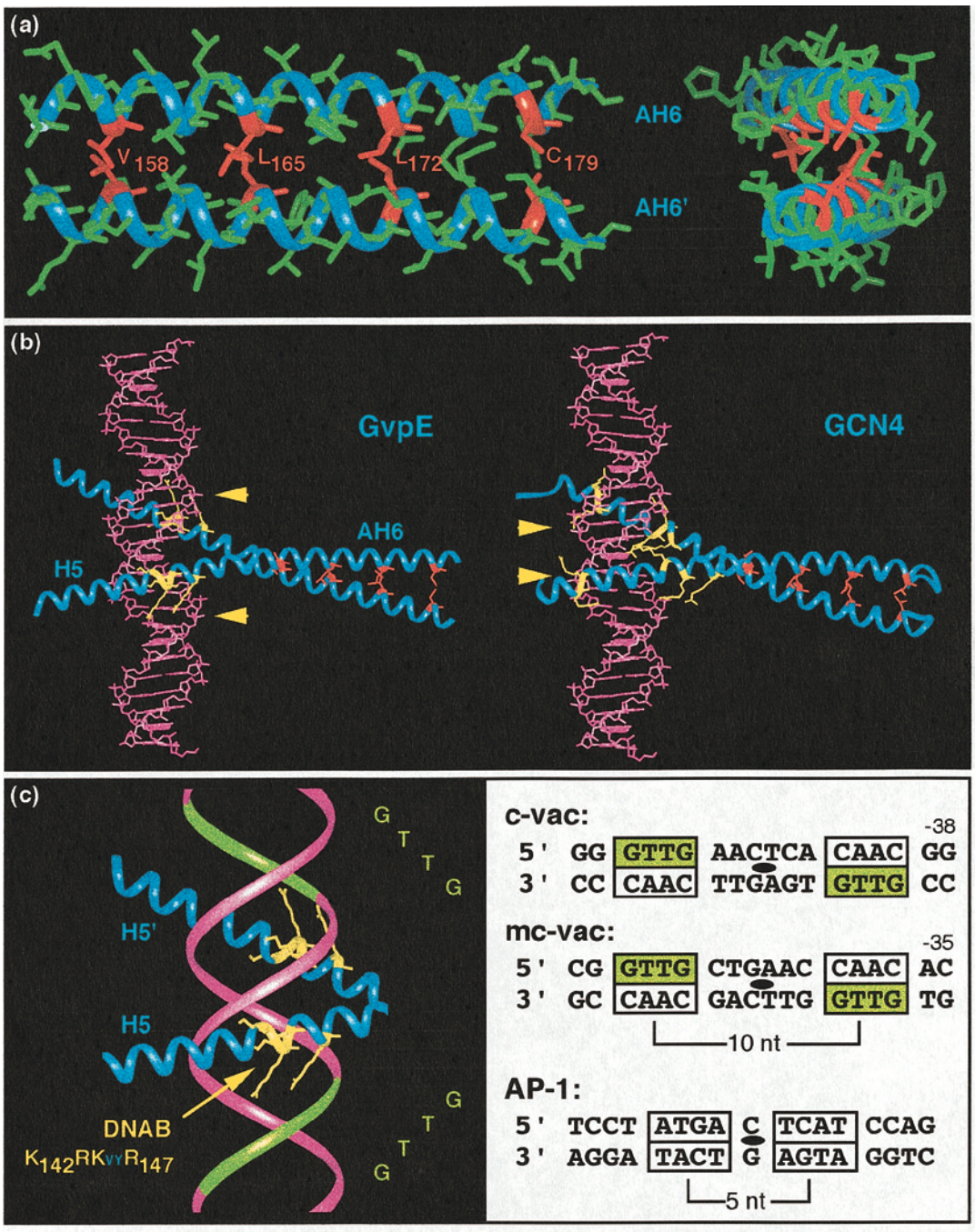

Figure 5. Molecular modelling of GvpE. (a) Model of the cGvpE leucine-zipper region Q152 to N182 showing two parallel oriented helices (AH6 and AH6′) within the putative cGvpE dimer. The hydrophobic core residues of the leucine-zipper are in red. On the right side, a view is shown along the axis of the coiled coil, emphasizing the localization of the hydrophobic core residues at the interface of the two helices. (b) Comparison of the modelled cGvpE bZIP dimer (H125-N182) docked to *B*-form DNA (left) and the crystal structure of the GCN4/DNA complex (right) (König & Richmond, 1993). The basic residues within DNAB are in yellow. The regions of the DNA backbone involved in salt-bridges with basic side-chains within DNAB are indicated by arrows. (c) Suggested interaction of the DNAB region of the modelled cGvpE dimer with DNA (left). Due to the distribution of basic residues N-terminal to the leucine-zipper region, amino acid residues of DNAB are probably interacting with DNA regions separated by one helix turn (green). Palindromic sequences, the half-sites of which are separated by exactly one helix turn, are found upstream from the GvpE-regulated *gvpA* promoters in both the c-vac and the mc-vac region (right, top). The eukaryal bZIP proteins recognize palindromic sequences arranged back to back on the DNA such as the AP-1 site (right, bottom).

form contacts with phosphate groups of a back-to-back palindromic sequence on opposite DNA strands within the same helix turn (König & Richmond, 1993), the GvpE model predicts that DNAB interacts with phosphate groups on opposite DNA strands that are separated by one helix turn (Figure 5(b)). This suggests that the putative GvpE bZIP dimer might bind to DNA by interacting with recognition half-sites that are located outside the scissor arms of the bZIP α-helices (Figure 5(c)). In contrast, GCN4 binds to DNA by recognizing half-sites between the helices (Ellenberger *et al.*, 1992). Support for our hypothesis for GvpE/DNA interaction comes from the finding that upstream of the *gvpA* promoter in the c-vac and mc-vac region (both of which are regulated by the respective GvpE) conserved palindromic sequences were found with half-sites separated by exactly one helix turn (Figure 5(c)).

Mutagenesis of the putative DNAB motif in GvpE, and effect on the activation of the c-*gvpA* expression

The three basic residues of the putative DNAB motif (K142, R143 and K144) were mutated by megaprimer PCR as described in Materials and Methods. The codons of these three amino acids were altered in c-*gvpE* to encode the amino acids AAA, ARA, ERA or ERE. The various mutated c-*gvpE* reading frames were inserted into the halobacterial expression vector pJAS35 and expressed under *fdx* promoter control in *Hf. volcanii*. Double transformants were constructed containing the wild-type, or a mutated c-*gvpE* in plasmid pJAS35, and the c-*gvpA* gene as target in vector pWL102 (A2 construct). In each transformant, the presence of both constructs was confirmed by Southern analysis (data not shown). In addition, a *Hf. volcanii* transformant was obtained containing the A2 construct itself.

Lysates of each transformant were subjected to Western analyses using (i) the anti-cGvpE serum (Krüger & Pfeifer, 1996) to analyze c-*gvpE* expression, and (ii) the anti-gas vesicle serum (Englert *et al.*, 1992b) to detect the cGvpA protein. A reaction with the anti-cGvpE serum was observed with each transformant containing a c-*gvpE* construct (Figure 6(a)), demonstrating that cGvpE as well as the mutated cGvpE proteins were formed in large amounts. Western analysis with the anti-gas vesicle serum gave no signal with the lysate of the A2 transformant, again underlining that the c-*gvpA* gene by itself is inactive (Figure 6(b)). The E/A transformant, however, contained the cGvpA protein, whereas none of the various Emut/A transformants synthesized cGvpA (Figure 6(b)). These results demonstrated that the basic amino acids at positions 142 to 144 in cGvpE are indeed required for cGvpE activity.

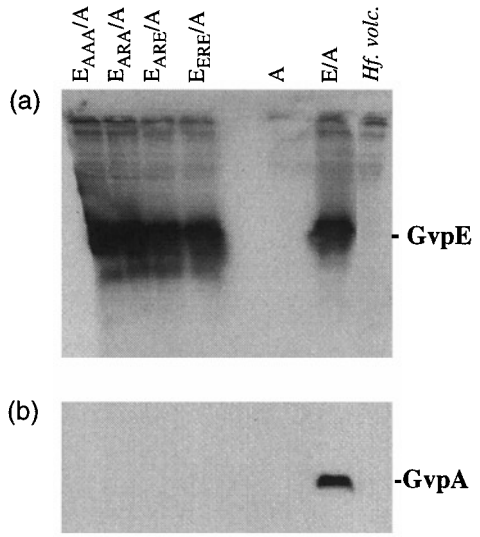

Figure 6. Western analysis of the *Hf. volcanii* transformants containing mutated c-*gvpE* reading frames and the respective controls. The analysis was done with 20 μg of protein each, and (a) anti-cGvpE serum (labelled GvpE) or (b) anti-gas vesicle serum (labelled GvpA) was applied.

Discussion

Hb. salinarum PHH4 produces gas vesicles exclusively in stationary growth phase, and their synthesis involves 14 *gvp* genes arranged in two clusters (Englert *et al.*, 1992a). The c-*gvpDE-FGHIJKLM* genes, located upstream of *gvpA* and oriented in the opposite direction, are cotranscribed at low level during all stages of growth, whereas the expression of the c-*gvpACNO* transcription unit encoding the two gas-vesicle structural proteins cGvpA and cGvpC is regulated at the transcript level (Krüger & Pfeifer, 1996). Transformation experiments demonstrated that the product of the c-*gvpE* gene is absolutely required for cA promoter activity, since the c-*gvpA* gene by itself is not transcribed in *Hf. volcanii* transformants. Similar results were found for the mc-vac region of *Hf. mediterranei*, where mcGvpE has been identified as the transcriptional activator for the mcA and mcD promoters of the mc-vac region (Röder & Pfeifer, 1996).

Closer inspection of the primary structure of the 21 kDa GvpE suggested that the protein resembles a basic leucine-zipper protein (bZIP), such as transcription factors characteristic for eukaryal gene regulation like cFos, cJun and GCN4 (Ellenberger, 1994). The secondary structure prediction implied several α-helices, including two helices (H5 and the amphiphilic helix AH6) in the C-terminal third of cGvpE that are separated by a cluster of basic amino acids. Similar to other bZIPs, the hydrophobic amino acid residues in the GvpE amphiphilic helix AH6 occur at distances of three and four

residues, and could form a hydrophobic interface between the two coiled helices of a putative active GvpE dimer. However, the arrangement of basic residues N-terminal to the predicted leucine-zipper differs from the constitution of the highly conserved DNAB region of eukaryal bZIP transcription factors such as GCN4 (Figure 4).

The significance of the basic amino acid cluster located N-terminal to helix AH6 for GvpE activity was tested by mutant analysis: the amino acid residues KRK (positions 142 to 144) were altered to residues with small or negatively charged side-chains (ARA, AAA, ERA or ERE). These GvpE mutants were studied *in vivo*, since *in vitro* DNA-binding studies are difficult to perform with halobacterial proteins that are adapted to 4 M KCl concentration. None of these cGvpE mutants was able to induce c-*gvpA* transcription in *Hf. volcanii* transformants, whereas the wild-type cGvpE could induce c-*gvpA* expression, demonstrating that this cluster of basic residues is absolutely required for cGvpE activity and possibly constitutes the DNA-binding site (DNAB). Mutations within the leucine-zipper part of the GvpE are also underway to test the hypothesis on GvpE dimerization in more detail.

Additional evidence for the putative DNAB motif at positions 142 to 144 in GvpE comes from the molecular model of the bZIP region (comprising residues H125 to N182 near the C terminus) docked to the major groove of *B*-form DNA, which revealed that the basic residues of DNAB could perfectly interact with the negatively charged DNA backbone. Furthermore, the model predicts that the DNAB interacts with phosphate groups of a palindromic sequence on opposite DNA strands, the half-sites of which are separated by one helix turn. This is different from other bZIP proteins, such as GCN4, which contacts phosphate groups of a back-to-back palindromic sequence on opposite DNA strands within the same helix turn (Ellenberger *et al*., 1992). A palindromic DNA sequence with half-sites separated by exactly one helix turn is located upstream of the cA (and mcA) promoter BoxA-element and could serve as GvpE binding site; this hypothesis is currently under investigation.

Interestingly, the predicted amphiphilic helix AH4, which is highly conserved among GvpE sequences of different organisms, also contains three basic residues (K103, R110 and R111) clustered on the hydrophilic side of AH4. This observation, along with the proximity of AH4 to the putative bZIP region, render it possible that AH4 might play a role in DNA major groove flanking the site in DNAB/DNA interaction.

The data presented here describe for the first time an archaeal gene regulator displaying striking similarity to eukaryal transcription factors. This is in accord with the finding that the basic transcription machinery and promoter sequences of archaea closely resemble their eukaryal counterparts (Reeve *et al*., 1997). The study of archaeal gene regulation is still in its infancy, but as more regulatory factors are identified, the more we will learn about the phylogenetic development of the eukaryal transcription machinery.

Materials and Methods

Constructs used for the transformation experiments

Fragments containing the cA, or pA promoter were amplified by polymerase chain reaction (PCR) using the synthetic oligonucleotides 5′ CTGGGTGGTCCATGGCCGCAACTACAG 3′ (containing a *Nco*I site, underlined) and 5′ CAAGATTTTGTACTGCAGAGCGAACTA 3′ (*Pst*I site underlined) for amplification of the 107 bp cA promoter fragment, and the oligonucleotides 5′ CTGGGATTACCATGGCTCTAACTATAC 3′ (*Nco*I underlined) and 5′ ACTCATTACACTGCAGATAACGACTGG 3′ (*Pst*I site underlined) for the amplification of the 85 bp pA promoter fragment. The PCR fragments were cut with *Pst*I and *Nco*I, and inserted upstream of the dihydrofolate reductase (*dhfr*) reading frame in vector pJAS33 conferring novobiocin resistance (M. Mevarech & J. Soppa, unpublished results). The *Pst*I site is located upstream of the original *dhfr* mRNA start site (Zusman *et al*., 1989), whereas the *Nco*I site overlaps the start codon of *dhfr* (see Figure 2, pJAS33).

Two different vectors containing the c-*gvpA* gene (constructs A1 and A2) were prepared. For construct A1, a 570 bp *Ssp*I/*Mlu*I fragment containing the c-*gvpA* gene and flanking sequences was cloned into the *Sma*I site of the halobacterial shuttle vector pMDS20 conferring novobiocin resistance (Holmes & Dyall-Smith, 1991). For construct A2 (c-*gvpA* in pWL102), this fragment was cloned into the *Sma*I site of the pBluescript II KS$^+$ vector (Stratagene, UA) after creating blunt ends in a filling up reaction with phage T4 DNA polymerase. The fragment was excised using the *Bam*HI and *Eco*RV sites of the pBluescript II KS$^+$ vector and ligated to the *Bam*HI and the refilled *Kpn*I site of the halobacterial vector pWL102 conferring mevinolin resistance (Lam & Doolittle, 1989). For the E-M construct, a 4.7 kb *Nru*I fragment containing the genes c-*gvpEFGHIJKLM* was blunt end-ligated into the *Sma*I site of the halobacterial vector pUBP2 (Blaseio & Pfeifer, 1990). In this construct, the c-*gvpE-M* genes are expressed from a fortuitous promoter located in the vector part (data not shown).

For the c-*gvpE* expression construct, the wild-type c-*gvpE* reading frame was amplified by PCR using the synthetic oligonucleotides 5′ CACGGAGATGGTGT*GGAT* CCATGGACGACCTC 3′ (=primer cE-*Nco*I; contains a *Bam*HI, printed in italics, and *Nco*I site, underlined) and 5′ GCCGTACGTGTAGAGGTAAGCTTCACTCATCC 3′ (=cE-*Hin*dIII). The PCR product was cut with *Bam*HI and *Hin*dIII, and ligated into the pBluescript II KS$^+$ vector. The c-*gvpE* reading frame was excised using *Nco*I and *Asp*718 and ligated to the *Nco*I- and *Asp*718-cut halobacterial expression vector pJAS35, where the reading frame inserted is expressed under *fdx* promoter control (Pfeifer *et al*., 1994).

Construction of the c-*gvpE* mutant genes by megaprimer PCR

Two mutation primers were constructed (cE-Mut1: 5′ AACTCACCgcGgcGgcGGTGTACCGCGTC 3′, and the degenerated primer cE-Mut2: 5′ AAATGAGCGA-

ACTCACCga/cGCGGga/cGGTGTACCGCGTC 3′; the mutations are indicated in lower-case letters). The sequence of the cE-Mut1 primer derives from positions 4010 to 4039 of the c-vac sequence (EMBL databank, accession number X94688), and encodes the amino acid residues AAA instead of KRK. The cE-Mut2 primer mixture (binding positions 4010 to 4047) encodes the amino acid residues ARA, ERA or ERE at positions 142 to 144 in the cGvpE sequence instead of KRK. Each one of these primers was used together with the primer cE-*Hin*dIII (see above) and *Hb. salinarum* PHH4 DNA as template in the first PCR resulting in a 200 bp fragment derived from sequences near the 3′ terminus of the c-*gvpE* gene. These 200 bp fragments containing the mutations were used as megaprimer in the second PCR to amplify the entire c-*gvpE* reading frame. For this reaction, the c-*gvpE* gene cloned in pBluescript II SK$^+$ served as template, and the megaprimer together with the M13 primer, binding near the 5′ terminus of c-*gvpE* in pBluescript, gave rise to the desired 700 bp c-*gvpE* fragments. The various fragments were cloned into pBluescript II SK$^+$ and analyzed by DNA sequence determination. Each of the desired mutations was found. The mutated c-*gvpE* fragments were excised using *Nco*I and *Asp*718, and ligated into the appropriately cut halobacterial expression vector pJAS35.

Transformation of *Haloferax volcanii* and Western analysis

Prior to transformation of *Hf. volcanii*, each construct was passaged through *E. coli dam*$^-$ strain GM1674 to avoid a halobacterial restriction system (Holmes *et al*., 1991). Transformation of *Hf. volcanii* was done as described (Horne *et al*., 1991; Englert *et al*., 1992a,b). Transformants were selected on agar plates containing 0.2 μg/ml novobiocin and/or 6 μg/ml mevinolin (or lovastatin). Lovastatin was a generous gift from MSD Sharp & Dohme GmbH (Haar, Germany). Total proteins of *Hf. volcanii* transformants were isolated from 5 ml cultures and tested by Western analyses as described (Krüger & Pfeifer, 1996).

Northern analysis and primer extension

The isolation of total RNA from *Hf. volcanii* transformants and Northern analysis was done as described (Krüger & Pfeifer, 1996). The DNA fragments used as probes were labelled according to the random priming procedure with [α-^{32}P]dATP (Feinberg & Vogelstein, 1983) using the 489 bp *Nco*I/*Asp*718 fragment containing the *dhfr* gene and the 236 bp *Kpn*I/*Dra*II fragment with the c-*gvpA* gene.

The determination of the transcriptional start site of the pJAS33-promoter constructs was done by primer extension analysis. The cDNA was produced as described by Reiter *et al*. (1988) with the Bethesda Research Laboratory reverse transcriptase kit, the (5′-^{32}P)-labelled oligonucleotide 5′ GTTCGCCGTCGCGGCCGATG 3′ (binds at nucleotides 61 to 42 of the *dhfr* gene (Zusman *et al*., 1989) and total RNA from *Hf. volcanii* and transformants. The labelled cDNA was analyzed on a 6% (w/v) polyacrylamide sequencing gel along with a sequencing ladder generated with the T3 primer (5′ ATTAACCCTCACTAAAG 3′) and pBluescript SK$^+$ DNA.

Sequence analysis and molecular modelling

The DNA sequence data of the entire c-vac region is deposited in the EMBL database under accession number X94688. Homology searching and alignment of the GvpE protein sequences was done with the FASTA algorithm (Pearson, 1990). Secondary structure prediction was performed using the neural network method of Rost & Sander (1993). Amphiphilic helices were spotted with an automated helical wheel projection procedure (Hermann, 1996) using the hydrophobicity indices due to Kyte & Doolittle (1982).

All three-dimensional modelling and visualisation was done with the SYBYL software package (Tripos, St. Louis). Homology modelling of the putative leucine-zipper region in the GvpE protein was performed manually on the basis of the backbone atomic coordinates of the homologous GCN4 protein, the crystal structure of which is known (Brookhaven Data Bank coordinate files 1YSA (Ellenberger *et al*., 1992) and 1DGC (König & Richmond, 1993). Sidechains of GvpE that were not identical with GCN4 were modelled by an algorithm implemented in SYBYL as described (Hermann *et al*., 1994). Docking of the modelled protein to DNA was performed manually based on the atomic coordinates of a crystallized GCN4/DNA complex (König & Richmond, 1993).

Acknowledgments

Part of this work was performed at the Max-Planck-Institut für Biochemie in Martinsried, Germany, and we thank Wolfram Zillig for hospitality and discussions. Hermann Heumann, Martin Lanzendörfer, Sonja Offner and Richard Röder are thanked for helpful suggestions, and Jobst Gmeiner and Kathryn Nixdorff for critical reading of the manuscript.

This work was financially supported by the Deutsche Forschungsgemeinschaft (Pf 165/6-2). T.H. acknowledges support from the Studienstiftung des deutschen Volkes and the European Molecular Biology Organization.

References

Blaseio, U. & Pfeifer, F. (1990). Transformation of *Halobacterium halobium*: development of vectors and investigation of gas vesicle synthesis. *Proc. Natl Acad. Sci. USA,* **87**, 6772–6776.

Danner, S. & Soppa, J. (1996). Characterization of the distal promoter element of halobacteria *in vivo* using saturation mutagenesis and selection. *Mol. Microbiol.* **19**, 1265–1276.

Dayhoff, M. O. (1978). A model of evolutionary change in proteins. Matrices for detecting distant relationships. In *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C., eds), vol. 5, suppl. 3, pp. 335–358.

Ellenberger, T. (1994). Getting a grip on DNA recognition: structures of the basic region leucine zipper,

and the basic region helix-loop-helix DNA-binding domain. *Curr. Opin. Struct. Biol.* **4**, 12–21.

Ellenberger, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted a helices: crystal structure of the protein-DNA complex. *Cell* **71**, 1223–1237.

Englert, C., Horne, M. & Pfeifer, F. (1990). Expression of the major gas vesicle protein gene in the halophilic archaebacterium *Haloferax mediterranei* is modulated by salt. *Mol. Gen. Genet.* **222**, 225–232.

Englert, C., Krüger, K., Offner, S. & Pfeifer, F. (1992a). Three different but related gene clusters encoding gas vesicles in halophilic archaea. *J. Mol. Biol.* **227**, 586–592.

Englert, C., Wanner, G. & Pfeifer, F. (1992b). 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 of its product. *Mol. Microbiol.* **6**, 3543–3550.

Feinberg, A. P. & Vogelstein, B. (1983). A technique for radiolabelling DNA restriction fragments to high specific activity. *Anal. Biochem.* **132**, 6–13.

Gropp, F., Gropp, R. & Betlach, M. C. (1995). Effects of upstream deletions on light- and oxygen-regulated *bacterio-opsin* gene expression in *Halobacterium halobium*. *Mol. Microbiol.* **16**, 357–364.

Hain, J., Reiter, W. D., Hüdepohl, U. & Zillig, W. (1992). Elements of an archaeal promoter defined by mutational analysis. *Nucl. Acids Res.* **20**, 5423–5428.

Hermann, T. (1996). Molekülmodellierung–ein Werkzeug der Molekularbiologie: Anwendung auf Nukleinsäuren und Nukleoproteinkomplexe. Dissertation, LMU Muenchen, Shaker Verlag, Aachen.

Hermann, T., Meier, T., Götte, M. & Heumann, H. (1994). The 'helix clamp' in HIV-1 reverse transcriptase: a new nucleic acid binding motif common in nucleic acid polymerases. *Nucl. Acids Res.* **22**, 4625–4633.

Holmes, M. & Dyall-Smith, M. (1991). Mutations in DNA gyrase result in novobiocin resistance in halophilic archaebacteria. *J. Bacteriol.* **173**, 642–648.

Holmes, M., Nuttal, S. & Dyall-Smith, M. (1991). Construction and use of halobacterial shuttle vectors for further studies on *Haloferax* DNA gyrase. *J. Bacteriol.* **173**, 3807–3813.

Horne, M. & Pfeifer, F. (1989). Expression of two gas vacuole protein genes in *Halobacterium halobium* and other related species. *Mol. Gen. Genet.* **218**, 437–444.

Horne, M., Englert, C., Wimmer, C. & Pfeifer, F. (1991). A DNA region of 9 kbp contains all genes necessary for gas-vesicle synthesis in halophilic archaebacteria. *Mol Microbiol.* **5**, 1159–1174.

Hu, J. C., O'Shea, E. K., Kim, P. S. & Sauer, R. T. (1990). Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science,* **250**, 1400–1403.

König, P. & Richmond, T. J. (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**, 139–154.

Krüger, K. & Pfeifer, F. (1996). Transcript analysis of the *c-vac* region and differential synthesis of the two regulatory gas vesicle proteins GvpD and GvpE in *Halobacterium salinarium* PHH4. *J. Bacteriol.* **178**, 4012–4019.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132.

Lam, W. L. & Doolittle, W. F. (1989). Shuttle vectors for the archaebacterium *Halobacterium volcanii*. *Proc. Natl Acad. Sci. USA,* **86**, 5478–5482.

Langer, D., Hain, J., Thuriaux, P. & Zillig, W. (1995). Transcription in Archaea: similarity to that in Eucarya. *Proc. Natl Acad. Sci. USA,* **92**, 5768–5772.

Lewit-Bentley, A., Morera, S., Huber, R. & Bodo, G. (1992). The effect of metal binding on the structure of annexin V and implications for membrane binding. *Eur. J. Biochem.* **210**, 73–77.

Offner, S. & Pfeifer, F. (1995). Complementation studies with the gas vesicle encoding p-vac region of *Halobacterium salinarium* PHH1 reveal a regulatory role for the p-*gvpDE* genes. *Mol. Microbiol.* **16**, 9–19.

Offner, S., Wanner, G. & Pfeifer, F. (1996). Functional studies of the p-*gvpACNO* operon of *Halobacterium salinarium* reveal that the GvpC protein shapes gas vesicles. *J. Bacteriol.* **178**, 2071–2978.

O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two–stranded parallel coiled coil. *Science,* **254**, 539–544.

Pabo, C. O. & Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095.

Palmer, J. R. & Daniels, C. J. (1995). *In vivo* definition of an archaeal promoter. *J. Bacteriol.* **177**, 1844–1849.

Pearson, W. R. (1990). Rapid and sensitive sequence comparison with FastP and FastA. *Methods Enzymol.* **183**, 63–98.

Pfeifer, F., Griffig, J. & Oesterhelt, D. (1993). The *fdx* gene encoding the [2Fe-2S] ferredoxin of *Halobacterium salinarium* (*H. halobium*). *Mol. Gen. Genet.* **239**, 66–71.

Pfeifer, F., Offner, S., Krüger, K., Ghahraman, P. & Englert, C. (1994). Transformation of halophilic archaea and investigation of gas vesicle synthesis. *Syst. Appl. Microbiol.* **16**, 569–577.

Qureshi, S., Bell, S. & Jackson, S. (1997). Factor requirements for transcription in the Archaeon *Sulfolobus shibatae*. *EMBO J.* **16**, 2927–2936.

Reeve, J. N., Sandman, K. & Daniels, C. J. (1997). Archaeal histones, nucleosomes, and transcription initiation. *Cell,* **89**, 999–1002.

Reiter, W. D., Palm, P. & Zillig, W. (1988). Transcription determination in the archaebacterium *Sulfolobus*: signal structures and linkage to transcription initiation. *Nucl. Acids Res.* **16**, 2454–2459.

Röder, 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.

Rost, B. & Sander, C. (1993). Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl Acad. Sci. USA,* **90**, 7558–7562.

Smith, P. D. & Moss, S. E. (1994). Structural evolution of the annexin supergene family. *Trends Genet.* **10**, 241–246.

Thomm, M. (1996). Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol. Rev.* **18**, 159–171.

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal of the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA*, **87**, 4576–4579.

Zusman, T., Rosenshine, I., Boehm, G., Jaenicke, R., Leskiw, B. & Mevarech, M. (1989). Dihydrofolate reductase of the extremely halophilic archaebacterium *Halobacterium volcanii*. *J. Biol. Chem.* **264**, 18878–18883.

Edited by K. Nagai

(Received 2 December 1997; received in revised form 12 March 1998; accepted 18 March 1998)