Divergent roles of CprK paralogues from *Desulfitobacterium hafniense* in activating gene expression

Krisztina Gábor,† Kagnew Hailesellasse Sene, Hauke Smidt,
Willem M. de Vos and John van der Oost

Wageningen University and Research Centre, Laboratory of Microbiology, Dreijenplein 10,
Wageningen 6703-HB, The Netherlands

Correspondence
Krisztina Gábor
krisztina.gabor@codexis.com

Gene duplication and horizontal gene transfer play an important role in the evolution of prokaryotic genomes. We have investigated the role of three CprK paralogues from the cAMP receptor protein–fumarate and nitrate reduction regulator (CRP–FNR) family of transcriptional regulators that are encoded in the genome of *Desulfitobacterium hafniense* DCB-2 and possibly regulate expression of genes involved in the energy-conserving terminal reduction of organohalides (halorespiration). The results from *in vivo* and *in vitro* promoter probe assays show that two regulators (CprK1 and CprK2) have an at least partially overlapping effector specificity, with preference for *ortho*-chlorophenols, while *meta*-chlorophenols proved to be effectors for CprK4. The presence of a potential transposase-encoding gene in the vicinity of the *cprK* genes indicates that their redundancy is probably caused by mobile genetic elements. The CprK paralogues activated transcription from promoters containing a 14 bp inverted repeat (dehalobox) that closely resembles the FNR-box. We found a strong negative correlation between the rate of transcriptional activation and the number of nucleotide changes from the optimal dehalobox sequence (TTAAT-N_4-ATTAA). Transcription was initiated by CprK4 from a promoter that is situated upstream of a gene encoding a methyl-accepting chemotaxis protein. This might be the first indication of taxis of an anaerobic bacterium to halogenated aromatic compounds.

INTRODUCTION

Members of the cAMP receptor protein–fumarate and nitrate reduction regulator (CRP–FNR) family of transcriptional regulators show exceptional diversity in their effector specificity and the promoters they target. As such, they control a wide range of physiological processes, ranging from nitrate and fumarate respiration (FNR) to glucose starvation by sensing cAMP levels (CRP) (Bauer *et al.*, 1999; Kolb *et al.*, 1993). At present, over 350 members of the CRP–FNR family have been identified, with many examples of more than one CRP–FNR homologue in the same organism (Korner *et al.*, 2003). The relationship between these related CRP–FNR transcriptional regulators is often complex, showing hierarchical characteristics. As an example, the facultative anaerobic γ-proteobacterium *Pseudomonas stutzeri* possesses four FNR homologues, of which FnrA, which contains the

N-terminal cysteine motif for iron–sulphur cluster binding, hierarchically controls one of the three DNR homologues that lack the cysteine-based sensory module (Vollack *et al.*, 1999). Another example of a regulatory cascade mediated by CRP–FNR homologues is found in *Rhodopseudomonas palustris* (Egland & Harwood, 2000). In this facultatively anaerobic α-proteobacterium, the degradation of 4-hydroxybenzoate is regulated by the CRP–FNR-type HbaR, which is hierarchically controlled by the oxygen-sensing AadR protein, also from the same family.

Due to the growing number of sequenced bacterial genomes, the known CRP–FNR family has expanded rapidly (Korner *et al.*, 2003). Genome analysis of *Escherichia coli* has identified the third member of the CRP–FNR family, YeiL (Anjum *et al.*, 2000). Apart from *Bradyrhizobium japonicum* and *Magnetospirillum magnetotacticum*, the chromosome of *Desulfitobacterium hafniense* DCB-2 shows the highest diversity of CRP–FNR-type regulators encoded in a single bacterium (Mesa *et al.*, 2006). *D. hafniense* belongs to the lineage of Gram-positive low-G + C bacteria, and its most important characteristic is the capability for anaerobic respiration with halogenated compounds as terminal electron acceptors (halorespiration) (Christiansen & Ahring, 1996). This strictly anaerobic

bacterium can couple the reductive dehalogenation of often toxic *meta-* and *ortho-*substituted phenol derivatives to its growth, thereby providing a potential means of bioremediation of polluted anoxic environmental sites (Van Eekert & Schraa, 2001). The key enzymes in halorespiration are the reductive dehalogenases (RDs), corrinoid/iron–sulphur-containing proteins that are predicted to be membrane anchored by a small protein, the gene for which generally clusters with the reductive dehalogenase-encoding gene (Smidt & de Vos, 2004). From *D. hafniense*, DCB-2, a halorespiration-inducible *ortho*-chlorophenol-reductive dehalogenase (CprA1) has been isolated and characterized (Christiansen *et al.*, 1998). The transcription of *cprA1* is activated by CprK1, a novel member of the CRP–FNR family, in the presence of 3-chloro-4-hydroxyphenylacetic acid (CHPA) (Gábor *et al.*, 2006). It has been shown that binding of CHPA to CprK1 results in an active DNA-binding conformation which enables the regulator to activate transcription from promoters that contain a specific DNA sequence termed the dehalobox (TTAAT-N_4-ATTAA) (Mazon *et al.*, 2007). The partially assembled sequence data for the genome of *D. hafniense* DCB-2 reveals the presence of at least 20 proteins that fulfil selection criteria for the CRP–FNR family, namely (i) a length of 230–250 aa, (ii) a C-terminal helix–turn–helix (HTH) DNA-binding motif, and (iii) an N-terminal nucleotide (effector)-binding domain (Korner *et al.*, 2003). Five of the CRP–FNR family members (including CprK1) show high similarity (36–89 % identity at the amino acid level) with CprK from *Desulfitobacterium dehalogenans* (Pop *et al.*, 2004) and cluster with genes that encode potential halorespiration proteins (Villemur *et al.*, 2002). We have focused on these putative transcriptional regulators in our research, while other CRP–FNR family members that do not cluster with halorespiration genes have been left for further investigations. The unusually high occurrence of CprK-like regulators is likely to be correlated with the relatively large number of halogenated compounds that this organism can accept as terminal electron acceptors (Madsen & Licht, 1992), enabling a specific response by each regulator to a specific group of halogenated compounds.

The aim of the present paper was to investigate the role of multiple CprK paralogues in *D. hafniense* by using *in vivo* promoter probe assays as well as *in vitro* DNA-binding assays. Our results indicate that at least two of the five CprK paralogues have a distinct effector range, which most likely reflects the gain of a new or specific function by divergence of redundant genes.

METHODS

Bioinformatics and statistical tools. The genome sequence of *D. hafniense* DCB-2 was screened for halorespiration genes using basic local alignment search tool (BLAST) (McGinnis & Madden, 2004). Protein accession numbers of the identified CprK homologues (based on *D. hafniense* genome sequence version 19-Jun-2006) and Dhaf locus tags from the IMG database (http://img.jgi.doe.gov), respect-

ively, are: ZP_01372871 and Dhaf_0678 (CprK1), ZP_01372893 and Dhaf_0698 (CprK2), ZP_01372888 and Dhaf_0693 (CprK3), ZP_01372914 and Dhaf_0718 (CprK4), and ZP_01372902 and Dhaf_0707 (CprK5). Conserved protein domains were identified with the help of the Pfam database (http://www.sanger.ac.uk/Software/Pfam), and HNN secondary structure prediction tool (http://npsa-pbil.ibcp.fr). The multiple sequence alignment was produced by CLUSTAL_X 1.81 (Thompson *et al.*, 1997) and edited in GeneDoc 2.6 (Nicholas *et al.*, 1997). The consensus dehalobox sequence was obtained using the WebLogo 3 BETA program (http://threeplusone.com/weblogo/) (Schneider & Stephens, 1990). The raw data output of the same program was used to obtain a weight measure (conservation value) for each position of the dehalobox consensus. The raw data output (given as bit/symbol) was converted to units of the graphical output data (bits) by multiplying the values by $1/\ln 2$. Correlation coefficients were computed in Excel (Microsoft Office 2003).

Identification of putative regulator-binding DNA sequences. The training datasets for the motif discovery MEME algorithm (Bailey & Elkan, 1994) consisted of 40 DNA sequences with an average size of 150 bp from the region directly upstream of the translational start codon of halorespiration genes. As a positive control, sequences that contained known FNR-like regulatory boxes from the *D. dehalogenans* chlorophenol-reductive dehalogenase (*cpr*) gene cluster were included in the analysis (Smidt *et al.*, 2000). Additionally, manual detection of inverted repeats was performed with the Palindrome tool (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=palindrome).

Overproduction and purification of CprK-like proteins. CprK1 was prepared as described previously (Gábor *et al.*, 2006). The genes encoding CprK2 and CprK4 were amplified from *D. hafniense* genomic DNA using oligonucleotide primers, as listed in Table 1. After digestion with the appropriate endonucleases, PCR products were cloned into linearized pET24d expression vectors (Novagen). The resulting T7-based expression vectors containing one of the *cprK* homologues (Table 2) were introduced into *E. coli* JM109(DE3) cells by heat-shock transformation. Overproduction of the recombinant proteins was done essentially as described for CprK1 (Gábor *et al.*, 2006), with the exception that IPTG-induced CprK2 and CprK4 production was carried out at 37 instead of 20 °C. Purification of CprK2 and CprK4 was performed similarly to that of CprK1, using sequential HiPrep heparin and Superdex 200 (Amersham Biosciences) chromatography columns (Gábor *et al.*, 2006).

***In vivo* promoter probe assays.** DNA fragments carrying potential promoter elements and CprK-binding sites (dehaloboxes) were PCR-amplified from *D. hafniense* genomic DNA and digested by endonucleases, followed by ligation with linearized pAK80 promoter probe vectors. The oligonucleotide primers and the resulting promoter probe vectors are listed in Tables 1 and 2, respectively. *E. coli* JM109(DE3) cells were co-transformed with the pAK80 derivatives and the corresponding expression vectors carrying a *cprK* homologue, as indicated: pWUR226 with pWUR216, pWUR218 or pWUR219; and pWUR227 with pWUR220, pWUR221, pWUR222 or pWUR223. Selection pressure during aerobic growth in Luria–Bertani medium was maintained by the addition of 30 μg kanamycin ml^{-1} (pET24d derivatives) and 200 μg erythromycin ml^{-1} (pAK80 derivatives). Triplicate experiments were carried out for each condition, and β-galactosidase activity was measured as described previously (Gábor *et al.*, 2006). Throughout the experiments, 1 Miller unit was defined as follows: $1000 \times A_{420}/(t \times v \times OD_{600})$, where A_{420} is absorbance at 420 nm, t is reaction time, v is sample volume and OD_{600} is the optical density of the culture at the time the sample was taken. In control experiments, recombinant cells carrying the pET24d vector instead of pWUR226 or pWUR227 were used.

Table 1. Oligonucleotide primers used in the study

Primer	Sequence*	Orientation†
BG1385	CGCGCCCATGGAAAGGGTAATAAGCAATCAC	S
BG1391	GCGCGCCATGGGAGAAATTCTTAAAAATTATATTTTTCC	S
BG1748	GGTTGAGAAATTCAGGTAAAG	S
BG1749	GGATCACATACGCAAGTATTAATG	A
BG1826	CGCGCAAGCTTGCTATGATGTTTATCTTCCTCC	S
BG1827	CGCGCGGATCCGGACGGCATCCTTCCTTTG	A
BG1828	GCGCGAAGCTTGGCCGGTCTTGTTGCCC	S
BG1830	CGCGCGGATCCGCTTCACCAGGAAAGGCGG	A
BG1831	GCGCGAAGCTTGCTCTTCTCGAAGAGGGGATAGC	S
BG1832	CGCGCGGATCCCAGGATAGTTCCCATTTTTTCACC	A
BG1835	GCGCGAAGCTTCCGAGGTTGAGAGCTTTAATTG	S
BG1836	CGCGCGGATCCGTACTCATCCCTTTCACCTCC	A
BG1838	GCGCGAAGCTTGGCCCCTAATTTATGGAG	S
BG1839	CGCGCGGATCCGCTAAAAACACAATCCCGGAC	A
BG1840	GCGCGAAGCTTCCTGAAGAGAGCCCTTGATC	S
BG1841	CGCGCGGATCCGGAATGAGCGGATCTTGAATG	A
BG1842	GCGCGAAGCTTGCAACCAGCGGCTTCGCC	S
BG1843	CGCGCGGATCCGATATGGATTGCACTAAGTTCCC	A
BG1940	GCGCAAGCTTAATAAGCTATCCCCTCTTCG	A
BG1941	GCGCGGATCCTCAGAATCTCAATTCCTCTTCC	A
BG2109	GCTCATTTTCCAAATTGGCG	S
BG2110	GGTACCAGAATAGTATAAAG	A

*Introduced endonuclease restriction sites are underlined.

†S, sense primer; A, antisense primer.

Electrophoretic mobility shift assay (EMSA). The 85 bp dsDNA fragment containing the DB8 dehalobox was PCR-amplified from *D. hafniense* genomic DNA using oligonucleotide primers BG2109 and BG2110. The 52 bp PCR product containing DB7 was obtained using BG1748 and BG1749 primers. DNA fragments were purified according to the modified 'crush and soak' method and 5′-labelled using [γ-^{32}P]ATP as described previously (Gábor *et al.*, 2006). EMSA reaction mixtures (20 µl) contained 1 × POP buffer (20 % glycerol, 50 mM Tris/HCl, pH 7.5, 5 mM MgCl$_2$, 2.5 mM EDTA, 250 mM NaCl), 2.5 mM DTT, 1 µg poly(dG-dC)-poly(dG-dC), 0.5 µM purified CprK1 or 1 µM purified CprK4, and 1.5–2 nM ^{32}P-labelled DNA. Additionally, potential effector molecules were added from a 2.7 mM aqueous stock solution, to a final effector to protein molar ratio of 800 : 1. The reaction mixtures were first incubated at 24 °C for 30 min to allow complex formation, then loaded on a 6 % polyacrylamide gel buffered with 89 mM Tris and 89 mM boric acid

Table 2. Plasmids used in this study

Plasmid	Description	Primers	Source or reference
pET24d	Expression vector (5.3 kb), pMB1 ori, Kan^R, IPTG-inducible T7 promoter	–	Novagen
pAK80	Promoter probe shuttle vector (11.0 kb), p15A/*L. lactis* ori, Ery^R, promoterless *lacLM* genes	–	Israelsen *et al.* (1995)
pWUR166	DB7:*lacLM* promoter fusion containing pAK80 derivative	BG1704/1743	Gábor *et al.* (2006)
pWUR168	DB6:*lacLM* promoter fusion containing pAK80 derivative	BG1702/1742	Gábor *et al.* (2006)
pWUR171	DB5:*lacLM* promoter fusion containing pAK80 derivative	BG1699/1782	Gábor *et al.* (2006)
pWUR176	*cprK1* gene cloned in pET24d	BG1379/1814	Gábor *et al.* (2006)
pWUR216	DB8:*lacLM* promoter fusion containing pAK80 derivative	BG1826/1827	This study
pWUR218	DB9:*lacLM* promoter fusion containing pAK80 derivative	BG1828/1830	This study
pWUR219	DB10:*lacLM* promoter fusion containing pAK80 derivative	BG1831/1832	This study
pWUR220	DB1:*lacLM* promoter fusion containing pAK80 derivative	BG1835/1836	This study
pWUR221	DB2:*lacLM* promoter fusion containing pAK80 derivative	BG1838/1839	This study
pWUR222	DB3:*lacLM* promoter fusion containing pAK80 derivative	BG1840/1841	This study
pWUR223	DB4:*lacLM* promoter fusion containing pAK80 derivative	BG1842/1843	This study
pWUR226	*cprK4* gene cloned in pET24d	BG1391/1940	This study
pWUR227	*cprK2* gene cloned in pET24d	BG1385/1941	This study

(~pH 8.1) and electrophoresed at 10 mA constant current at 4 °C, followed by drying and autoradiography. Band intensities were quantified with GeneTool 3.08 (Syngene) software using the manual baseline correction method.

RESULTS

CprK-like regulators and target sequences in *D. hafniense*

The genome of the halorespiring *D. hafniense* DCB-2 contains five ORFs (*cprK1*–*cprK5*) that share 36–89 % amino acid sequence identity with the transcriptional regulator CprK from *D. dehalogenans* (Smidt *et al.*, 2000). A BLAST search also identified several clustering genes that, similarly to the previously annotated *cprTKZEBACD* cluster of *D. dehalogenans* (Smidt *et al.*, 2000), were predicted to encode the following: chlorophenol reductive dehalogenases (*cprA*), small hydrophobic membrane anchors for the reductive dehalogenase (*cprB*), GroEL-type chaperones (*cprD* and *cprE*), proline *cis/trans* isomerases (*cprT*), and putative transcriptional regulators from the NosR/NirI family (*cprC*) (Fig. 1a). One of the gene clusters contains a putative transposase-encoding gene from the IS4 family (*trn*) that is next to a truncated reductive dehalogenase gene (*rdhΔ*) and a Tat-signal peptide-encoding ORF (*tat*). All the five *D. hafniense* CprK paralogues were predicted to contain an N-terminal effector-binding domain (Pfam no. PF00027; cNMP-binding domain) and a conserved C-terminal HTH DNA-binding motif (Pfam no. PF00325). Indeed, the recently solved crystal structure of CprK1 from *D. hafniense* confirmed the presence of these domains (Joyce *et al.*, 2006).

In the recognition α-helixes (α_F) of the HTH domain, CprK1 and CprK2 possess a conserved motif V--SR (Fig. 2); in CprK4 the corresponding sequence is V--SK, whereas in CprK3 and CprK5 it is V--CK. The position of these residues corresponds to the motif E--SR in *E. coli* FNR, which has been demonstrated to be responsible for the specificity towards its target DNA sequence TTGAT-N_4-ATCAA (FNR-box) (Green *et al.*, 2001).

Based on the high similarity of the CprK paralogues to characterized CprK transcriptional regulators as well as on their chromosomal position, a search was performed for regulator-binding motifs in the upstream sequences of all the *cpr* genes. Ten putative regulatory sequences, termed dehaloboxes (DB1–10), were identified using the MEME algorithm (Bailey & Elkan, 1994); in addition, one distantly related motif was found in the promoter region of the putative 3,5-dichlorophenol (3,5-DCP) reductive dehalogenase-encoding gene *cprBA5* (DB0) (Fig. 1a, b). The identified sequences included three dehaloboxes that have been studied in detail previously (Gábor *et al.*, 2006). Each dehalobox contains a 5 bp imperfect inverted repeat with 4 nt spacing, and the consensus sequence of dehaloboxes is TTAGT-N_4-ACTAA (Fig. 1b).

Transcriptional activation by CprK paralogues in *E. coli*

In order to study the nature of the putative CprK-binding motifs (dehaloboxes) that were found upstream of the *cpr* genes (Fig. 1a, b), a recombinant *in vivo* promoter probe assay was developed in *E. coli*. Each *cprK*-encoding gene was PCR-amplified from *D. hafniense* genomic DNA, cloned into the pET24d expression vector and used to transform *E. coli* JM109(DE3) host cells. After optimization of the culture conditions for protein synthesis, CprK1 (26.5 kDa), CprK2 (26.5 kDa) and CprK4 (27 kDa) were successfully overproduced in *E. coli*, partly in the soluble fraction (Fig. 1c). Despite repeated efforts, the soluble overproduction of CprK3 and CprK5 was not achieved. CprK1 appeared as a double band during SDS-PAGE analysis (Fig. 1c); MS measurements have demonstrated that this is (at least in part) due to removal of the N-terminal methionine (Mazon *et al.*, 2007). A promoterless β-galactosidase-encoding gene was fused with *D. hafniense* promoter fragments, each containing a dehalobox centred in the middle of a 0.2 kb DNA fragment. The resulting promoter probe constructs (pWUR216–223) are listed in Table 2.

We tested the *in vivo* transcriptional activation within gene clusters by co-transforming *E. coli* JM109(DE3) host cells with the CprK2-encoding plasmid (pWUR227) and dehaloboxes from its own gene cluster (DB1–DB4), or the CprK4-encoding plasmid (pWUR226) and dehaloboxes DB8–DB10. Triplicate cultures of each promoter fusion-containing strain were grown under two conditions: in the presence of 20 mM CHPA as effector compound, or in the absence of the halogenated compound. The functionality of CprK2 and CprK4 on the respective promoters was analysed by detecting the activity of the β-galactosidase reporter enzyme (Fig. 3), and data were compared with results that had previously been obtained for CprK1 on DB5–DB7 (Gábor *et al.*, 2006).

CprK2 was most active on the promoter that contains dehalobox DB1 and is situated upstream of the putative reductive dehalogenase-encoding genes *cprBA2* (Fig. 3a). Similar activity was measured on the promoter of *cprK3*, containing DB3. The putative promoters that contain DB2 and DB4 dehaloboxes were weaker, showing approximately 40 % residual activity.

Remarkably, transcriptional activation by CprK4 was only observed with DB8, a dehalobox that is situated upstream of a putative methyl-accepting protein-encoding gene (*macA*) (Fig. 3c). The fact that the promoter of *macA* encoding an important player in the chemotaxis machinery is recognized by a CprK homologue might suggest that *D. hafniense* elicits chemotactic responses to chlorinated compounds. Promoter activity mediated by CprK4 from DB9 or DB10 (upstream of *cprK4* and *cprBA4*, respectively) was negligible. This indicates that the transcription of *cprK4* and *cprBA4* might be activated by a different regulator, or alternatively, that their transcription can be
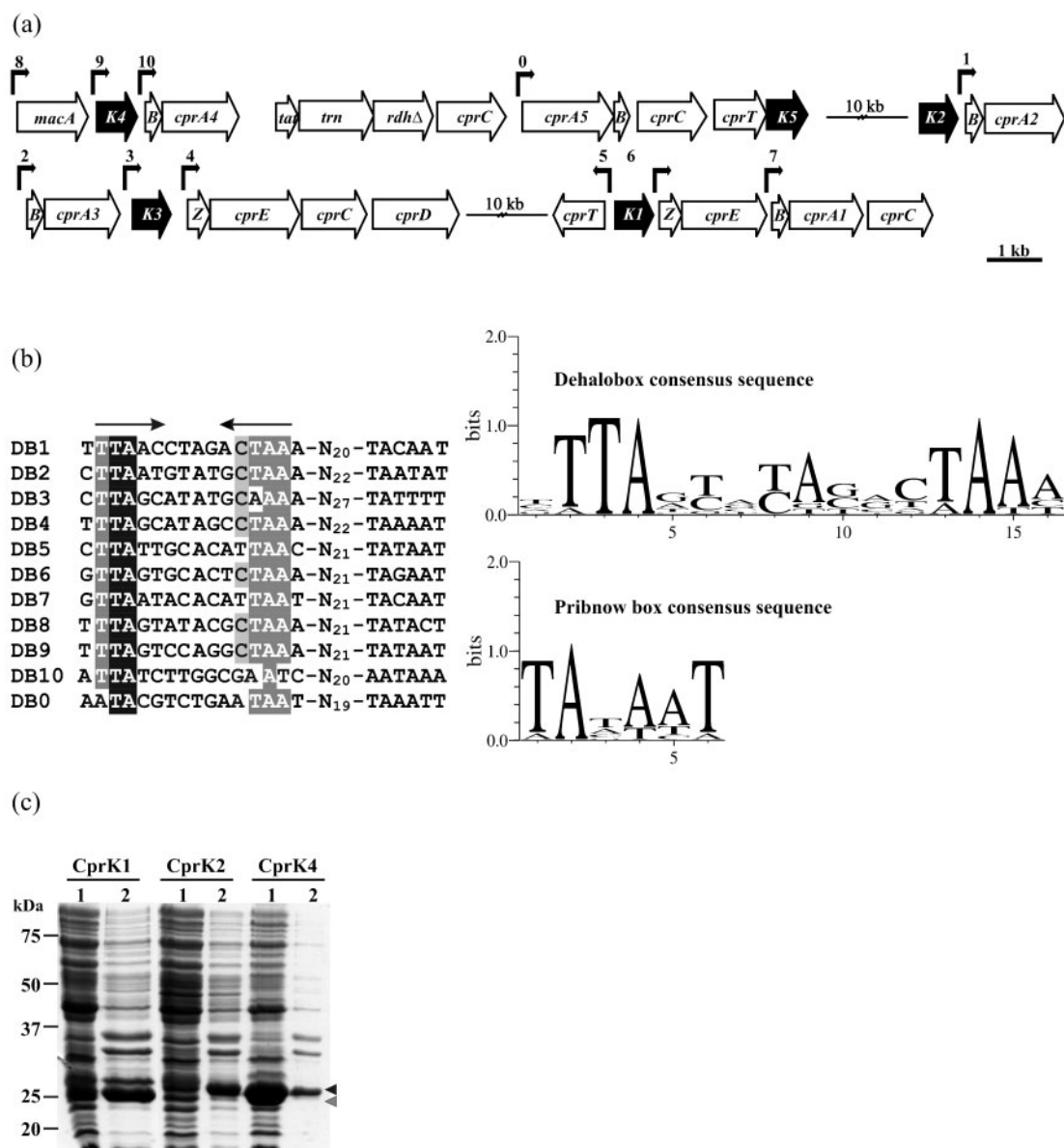
Fig. 1. Identification and production of CprK homologues. (a) Reductive dehalogenase gene clusters in *D. hafniense* DCB-2. Arrows indicate putative CprK-binding sites (dehaloboxes). (b) Sequence alignment of dehaloboxes. Arrows are placed above the 5 bp inverted repeats; the 3′ nucleotides correspond to putative Pribnow boxes. Dehalobox and Pribnow box consensus sequences generated by the WebLogo 3 BETA program are shown to the right of the alignment. (c) SDS-PAGE gel of the soluble (1) and insoluble (2) cell fractions of recombinant *E. coli* JM109(DE3) cells overproducing CprK1 (26.5 kDa), CprK2 (26.5 kDa) or CprK4 (27 kDa). Full-size heterologously produced proteins are indicated by the black arrowhead; the grey arrowhead corresponds to the truncated CprK1 protein.

initiated by CprK4 from *macA* by readthrough. One important difference was observed in the pattern of transcriptional activation by CprK-like proteins: while CprK1 (Fig. 3b) and CprK2 (Fig. 3a) were strongly regulated by CHPA, showing a drastically reduced promoter activity when this compound was absent from the media, CprK4 was equally active with or without CHPA (Fig. 3c). We have studied this phenomenon further using *in vitro* EMSA experiments.

## Dehaloboxes: perfect fit on a perfect repeat?

In vivo promoter probe experiments showed that, similarly to CprK1, the two new CprK paralogues (CprK2 and CprK4) are also involved in the recruitment of RNA polymerase (RNAP) and consequently in transcriptional activation in *E. coli* (Fig. 3). All the *D. hafniense* promoter fragments that were used in the assay harbour a 14 bp inverted repeat, termed the dehalobox, as a potential
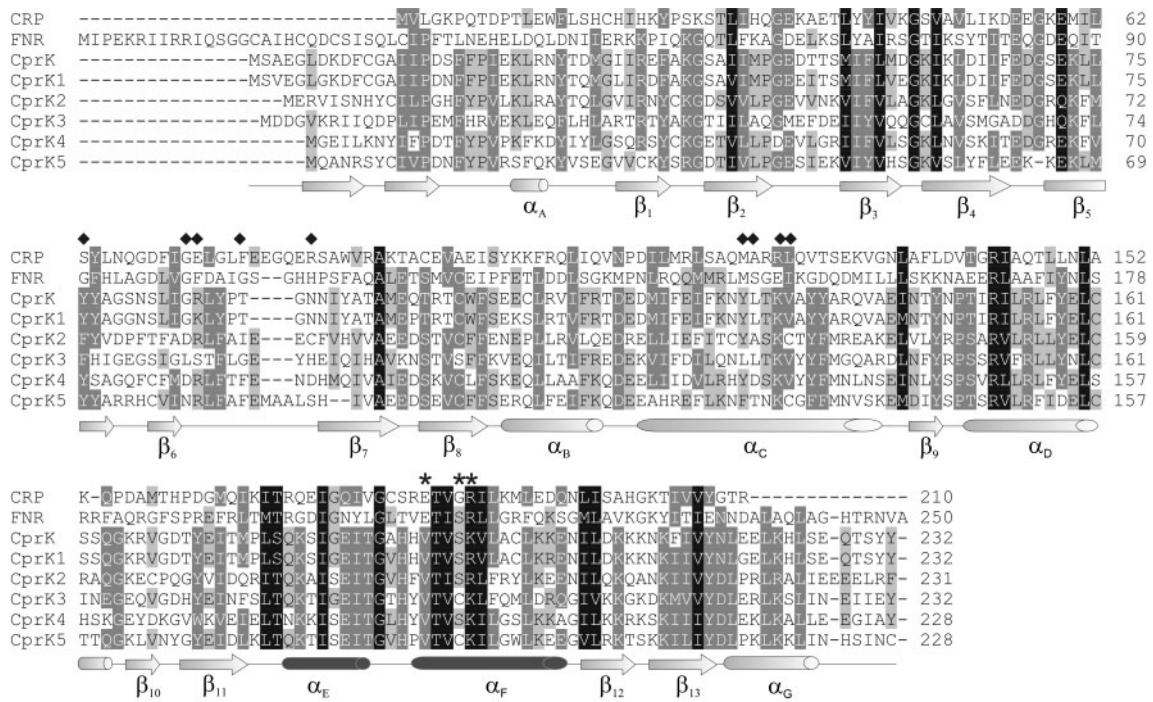
Fig. 2. Multiple alignment of the protein sequence of selected members of the CRP–FNR family. The secondary structure of CprK1 is included below the alignment (based on PDB 2H6B, monomer A); α-helices are denoted by cylinders, β-sheets by arrows. The cylinders corresponding to the characteristic HTH DNA-binding motif are in black. Residues that are involved in CHPA binding in CprK1 are indicated by diamonds, while asterisks highlight amino acids that confer DNA-binding specificity to FNR and to the CprK homologues. CRP, cAMP-binding protein from *E. coli*; FNR, fumarate and nitrate respiration regulator from *E. coli*; CprK, transcriptional regulator of halorespiration from *D. dehalogenans*; CprK1–K5, known or putative transcriptional regulators of halorespiration from *D. hafniense*.

binding target for the CprK paralogues. CHPA-induced promoter activities varied by two orders of magnitude, while only a two- to threefold variation was observed in the yield of functional (i.e. soluble) production of CprK homologues by *E. coli* JM109(DE3) (Fig. 1c). Hence, the difference in protein concentration does not explain the observed 100-fold difference in the promoter activities. Among the 10 studied dehaloboxes, only DB7 is a perfect inverted repeat (TTAAT-N_4-ATTAA). In addition, it is notable that among all the promoters, that of *cprBA1*
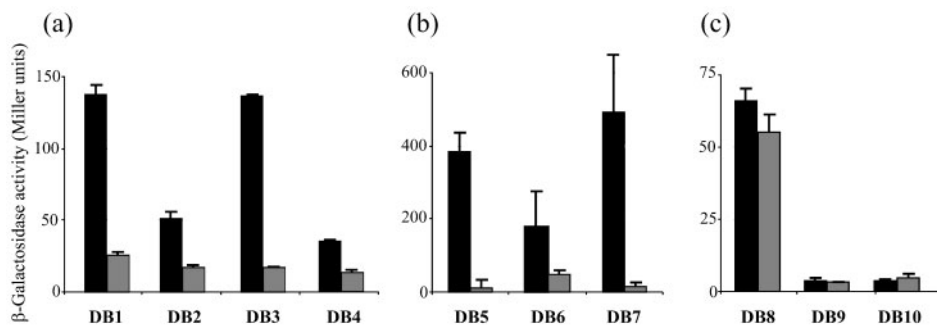
Fig. 3. *In vivo* promoter probe assays using recombinant *E. coli* JM109(DE3) cells overproducing CprK2 (a), CprK1 (b) or CprK4 (c). Dehaloboxes (DB1–10) centred in *D. hafniense* promoter fragments were fused to β-galactosidase-encoding genes and tested for transcriptional activation in the presence of 20 mM CHPA (black bars) or in the absence of the effector (grey bars). Results shown in (b) were obtained previously (Gábor *et al.*, 2006) and are included for comparison. Error bars, SDs for mean values from three biological replicates.

containing DB7 shows the highest promoter activity, not only in the presence of CprK1 (500 Miller units; Fig. 3b) but also with CprK4 (130 Miller units; results not shown). This makes DB7 a candidate optimal binding target for CprK-like proteins. We investigated the correlation between the promoter activity and changes from the optimal target sequence (DB7, gTTAATacacATTAAt). First, for each nucleotide position of the dehaloboxes, a weight measure (conservation value) was assigned using the Weblogo 3 BETA program. A high value indicated a highly conserved position, while a low number suggested that there is no preferred nucleotide at a certain position. Next, we compared each dehalobox to DB7. At the positions that showed discrepancies from the optimal binding sequence, the sums of the weight measure values were taken as an indication of the degree of discrepancy. Finally, we computed the correlation between the degree of discrepancy and the corresponding promoter activity. Correlation was measured on a 0–1 scale, where values up to 0.4 show negligible to weak correlation, while values above 0.7 indicate a strong to very strong correlation between two variables. A strong negative correlation (-0.76) was observed between the measured reporter enzyme activity and the degree of discrepancy from the putative optimal target sequence TTAAT-N_4-ATTAA (Fig. 4). Similar results were obtained (correlation coefficient was -0.68) when changes in the consensus Pribnow sequence (TATAAT) were also taken into account.
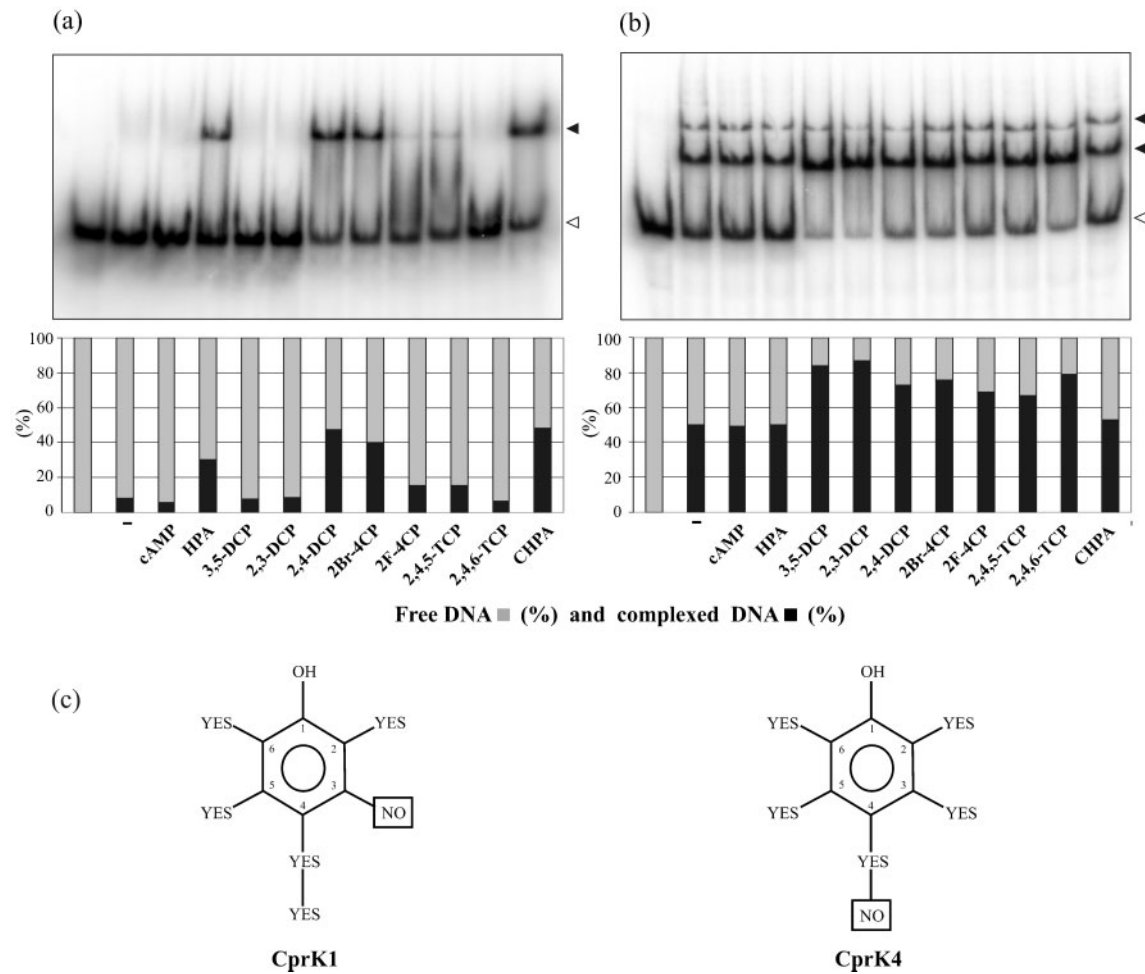
Alternative effectors of CprK1 and CprK4

In vivo promoter probe experiments in *E. coli* revealed an unusual feature of CprK4: although it distinguished between preferred (DB8) and non-preferred (DB9 and DB10) *D. hafniense* promoter fragments, it was constitutively active, irrespective of the presence or absence of the potential effector CHPA (Fig. 3c). There are two possible explanations for this behaviour: either CprK4 was activated by cAMP (a molecule that is present in *E. coli* but cannot

be synthesized in *D. hafniense*); or alternatively, the effector-free CprK4 is already capable of DNA binding and this ability is only enhanced in the presence of a true ligand molecule (apparently not by CHPA). In order to study the DNA-binding properties of CprK4 in the presence of different potential effector molecules, the protein was purified and EMSAs were performed. This *in vitro* method overcomes the difficulties raised by the toxic nature of these compounds when used in growth experiments. In parallel, the alternative effector range of CprK1 was mapped with the same method. Attempts to functionally purify CprK2 were unsuccessful, due to the instability of the protein; therefore, CprK2 was excluded from the EMSA experiments. Binding experiments were performed with small promoter fragments containing dehaloboxes for which the highest *in vivo* activity was measured previously (DB7 for CprK1 and DB8 for CprK4; Fig. 3). Labelled DNA fragments (1.5–2 nM) were incubated in the presence of 1 μM purified regulator protein and 800 μM potential effector molecules. These chemicals included cAMP, 4-hydroxyphenylacetic acid (HPA) (a dechlorinated derivative of CHPA), di- and trichlorinated phenolic compounds that have been reported to be degraded by *D. hafniense* DCB-2 and/or the closely related *D. dehalogenans* [3,5-DCP, 2,3-DCP, 2,4-DCP, 2,4,5-trichlorophenol (TCP) and 2,4,6-TCP] (Christiansen & Ahring, 1996; Utkin *et al.*, 1995), and CHPA and its bromo- and fluoro-substituted derivatives (2-Br-4-CP and 2-F-4-CP). The experiments revealed that CprK4 is indeed capable of DNA binding in its effector-free form (Fig. 5b), and that cAMP or CHPA do not promote DNA binding to any extent. However, in the presence of compounds such as 3,5-DCP, 2,3-DCP and 2,4,6-TCP, over 80 % of the DNA formed a complex with CprK4. It is interesting to note that exactly these compounds had no potential to act as effectors for CprK1 (Fig. 5a). This suggests a complementary function of the two paralogues. The analogues 2,4-DCP and CHPA are very good effectors for CprK1, similarly to 2-Br-4-CP. The related compound that contained a fluorine atom (2-F-4-CP) was a weak effector, similarly to 2,4,5-TCP. Unexpectedly, in the presence of a high molar excess of HPA, protein–DNA complexes were also formed. In the case of CprK4, addition of 2,4-DCP, 2-Br/F-4-CP and 2,4,5-TCP resulted in an intermediate level of increase in protein–DNA complex formation (70–75 % complexed from the total DNA). In summary, the results indicate that a halogen substitution at the *meta* position in phenol derivatives resulted in the loss of effector activity on CprK1; in contrast, this appeared to be beneficial for the activity of CprK4 (Fig. 5c). The only restriction for CprK4 is that it cannot accept a long side-chain (such as the acetic acid tail in CHPA) at the *para* position, probably due to steric constraints in the effector-binding pocket. These results are in agreement with those obtained by micro electrospray ionization MS (μESI-MS) (Mazon *et al.*, 2007).

The last notable difference between the *in vitro* DNA-binding properties of the two proteins is the formation of a
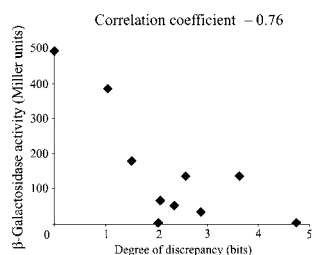
Fig. 4. Correlation between promoter activity mediated by the CprK paralogues and the degree of discrepancy from the optimal regulator-binding sequence. Promoter activity is expressed in Miller units (enzymic activity of the reporter protein); the degree of discrepancy is calculated as the sum of the weight measure values of each nucleotide position that differs from the optimal dehalobox sequence DB7 (gTTAATacacATTAAt).

Fig. 5. Alternative effectors of CprK1 and CprK4. EMSA was performed in the presence of halogenated compounds and compounds with related structure (as indicated below the image), and with constant amounts of CprK1 (a) or CprK4 (b). The reaction mixture contained dehalobox DB7 (CprK1) or DB8 (CprK4) and 800 : 1 molar ratios of effector and protein. The first lane on the left in each gel served as a control, containing only labelled DNA. Free DNA is indicated by open arrowheads, protein–DNA complexes are indicated by filled arrowheads. Free DNA or DNA in complex with the regulator was quantified as a percentage of the total DNA loaded in each well. (c) The positions in phenol derivatives where halogen or other substitutions can or cannot allow DNA–protein complex formation by CprK1 or CprK4.

second, slower-migrating protein–DNA complex with CprK4, which is not observed with CprK1 (Fig. 5b). This indicates that more CprK4 molecules are bound to the same DNA fragment, causing a further decrease in its mobility, either by forming a higher CprK4 oligomer (most likely a tetramer, as proposed by Joyce et al., 2006), which then binds to the DNA, or by the binding of a second CprK4 dimer to a weak, low-affinity DNA site. One weak inverted repeat (TTATC-N_4-ATTTA) that resembles the dehalobox consensus was indeed identified at a position that overlaps DB8. The possible, although fairly unlikely, functionality of this putative binding site could give an explanation of how CprK4-mediated transcriptional activation can be controlled by a co-operative action of a second transcriptional regulator.

DISCUSSION

Whole-genome shotgun sequencing of the *D. hafniense* DCB-2 genome revealed a redundancy in genes encoding potential reductive dehalogenases (*cprA*) and putative CRP–FNR-type transcriptional regulators (*cprK*), which enables this organism to gain energy from halorespiration (Smidt & de Vos, 2004; Villemur et al., 2002). Moreover, these *cprA*- and *cprK*-like genes were found to be located in a conserved gene context resembling those found in other halorespiring bacteria (Smidt et al., 2000). The work presented here focused on the diversity of predicted CprK regulators and aimed to gain more understanding of the roles of multiple CprK homologues in strain DCB-2 (overlapping or complementary function). In this

discussion we also suggest how their specificity is exerted, or in other words, how cross-talk of these highly similar regulators is eliminated, and what is the origin of the redundancy of the CprK-like proteins.

We found that the CprK paralogues in *D. hafniense* DCB-2 have at least partly overlapping functions, as both CprK1 and CprK2 initiated transcription from dehalobox-containing promoters in the presence of CHPA. CprK1 and CprK2 are 62 % similar at the amino acid level, which might indicate that the two regulators can substitute for each other, and that they both control regulons with identical, or at least overlapping, functionality. Such interchangeable function has been demonstrated in *Lactococcus lactis*, in which two closely related paralogues from the CRP–FNR family are present: FlpA and FlpB (75 % similarity), encoded by distinct operons involved in metal-ion transport. The single *L. lactis* Flp mutants behave similarly to the wild-type cells; only the double mutant shows an altered polypeptide profile and drastically decreased zinc content (Gostick *et al.*, 1999). We also found that one CprK paralogue (CprK4) responds to a different set of effector molecules, indicating that it has a function complementary to that of CprK1. Specific DNA binding mediated by CprK4 was enhanced in the presence of *meta*-substituted compounds such as 3,5-DCP and 2,3-DCP. In contrast, these compounds did not induce DNA binding by CprK1. The recently elucidated crystal structure of CprK1 reveals that such differences in effector specificity are expected, since among the CHPA-binding residues of CprK1, only Lys133 is conserved in all the CprK paralogues (CprK4 shares five identical effector-binding amino acids with CprK1) (Fig. 2) (Joyce *et al.*, 2006). In conclusion, we showed that CprK2 and CprK4 can sense a range of halogenated phenol derivatives in a co-operative manner, with only partly overlapping specificity.

The large number of transcriptional regulators from the CRP–FNR family in the same organism raises the question of how target genes are discriminated in the simultaneous presence of these regulators. It has been speculated that subtle changes in the target DNA sequence, the exertion of hierarchical control and further protein requirements (co-activators and specific sigma factors) ensure the elimination of unwanted cross-talk of the similar regulators (Zumft, 1997). We showed that a strong negative correlation exists between deviations from the optimal CprK-binding sequence (TTAAT-N_4-ATTAA) and the promoter activity determined by using *lacZ* promoter fusions in *E. coli*. Conservation of the A/T and T/A base pairs at the symmetry-related positions 3 and 12 in dehaloboxes ensures that these promoters are not recognized by proteins with the FNR-type E–SR motif in their recognition α-helix (Gábor *et al.*, 2006). It has also been demonstrated that both the spacing nucleotides of the inverted repeat and the nucleotides further downstream affect transcriptional activation, which enables fine tuning of promoter activity (Scott *et al.*, 2003; Veldman *et al.*, 2006). As CprK4 showed a constitutive DNA-binding

activity, it is possible that in *D. hafniense* the expression of *cprK4* is regulated in a hierarchical way, as described for FixK in *Rhizobium meliloti* (Batut *et al.*, 1989). The expression of *fixK* is activated in an oxygen-dependent manner by the two-component signal transduction proteins FixLJ. Also, hierarchy can be exerted by means of other regulatory interactions. As an example, in *Paracoccus denitrificans* the role of three FNR homologues (FnrP, NNR and NarR) is separated by the need for the presence of a specific sigma factor (Van Spanning *et al.*, 1997; Veldman *et al.*, 2006). However, in the CprK paralogues most of the characteristic negatively charged residues that make contact with the common σ^{70} factor of the *D. hafniense* RNAP holoenzyme are conserved; thus, it is unlikely that they utilize an alternative sigma factor (Gábor *et al.*, 2006). The presence of co-activators or repressors can also alter promoter activity (Barnard *et al.*, 2004; Bearson *et al.*, 2002). The results from EMSA experiments indicate that the promoter regulated by CprK4 might contain one additional, weak-affinity regulator-binding site, as shown by the formation of a second, slower-migrating band (Fig. 5b). This distinct complex might have deviating features, e.g. serving as the target for a co-regulator. The third possibility is that the physical shape of the DNA in complex with CprK4 is different under effector-free and effector-bound conditions, as has been shown for LysR-type regulators (Diaz & Prieto, 2000), which would enable RNAP recruitment only under halorespiring conditions in *D. hafniense*.

Finally, we have demonstrated that at least three CprK paralogues are involved in sensing a range of halogenated compounds in *D. hafniense*. Genomic analyses have revealed that a large number of gene paralogues may be present in bacterial chromosomes. It is known that gene duplications and horizontal gene transfer (HGT), which give rise to gene paralogues, play an important role in the evolution of bacterial genomes by providing a rapid means of adaptation to environmental changes (Hurles, 2004; Janssen *et al.*, 2005; Van der Meer & Sentchilo, 2003). An interesting theory of Parales & Harwood (2002) suggests that chemotaxis provides an excellent opportunity to promote HGT. It directs motile bacteria towards contaminated sites where strains carrying the relevant catabolic plasmids are likely to be present and therefore can donate new catabolic genes to the recipient chemotactic bacteria. Remarkably, our results indicate that the promoter of *macA*, a putative methyl-accepting chemotaxis protein-encoding gene, is recognized by CprK4. *D. hafniense* DCB-2 possesses one or two terminal flagella (Christiansen & Ahring, 1996), and with one exception (the phosphoprotein phosphatase *cheZ*) all the *che* genes required for the chemotaxis machinery are present in the genome: the core receptor-encoding *cheA*, *cheY* and *cheW*, and the modulating *cheR* and *cheB* (Webre *et al.*, 2004); therefore, it is tempting to assume that this might enable the bacterium to detect and respond to specific chemicals (potentially including organohalides) in the environment. Hence, it is

possible that this particular CprK homologue regulates not only genes that are involved in the degradation of haloaromatic compounds but also a gene that enables the organism to actively seek these compounds. Such a relationship has been described in the 4-hydroxybenzoate (4-HB)-degrading *Pseudomonas putida* PRS2000 strain between the PcaR transcriptional regulator and PcaK, which is involved in the chemotaxtic response to 4-HB (Harwood *et al.*, 1994), and for NahR and NahY from *P. putida* G7 for the degradation of, and chemotaxis to the aromatic hydrocarbon naphthalene (Grimm & Harwood, 1999; Schell, 1985).

In conclusion, we have investigated the role of multiple CprK paralogues in *D. hafniense* DCB-2, and found that at least two regulators (CprK2 and CprK4) have a distinct effector-sensing function. The DNA-binding specificity is possibly exerted by changes in the target nucleotide sequence from the consensus dehalobox sequence. The redundancy of the *cprK* genes is likely to be caused by mobile genetic elements, as shown by the presence of a putative transposase-encoding gene in the vicinity of the *cpr* gene clusters. The presence of a halorespiration-inducible methyl-accepting chemotaxis protein-encoding gene in *D. hafniense* suggests that chemotaxis plays an important role in actively seeking halogenated compounds.

ACKNOWLEDGEMENTS

REFERENCES

Anjum, M. F., Green, J. & Guest, J. R. (2000). YeiL, the third member of the CRP–FNR family in *Escherichia coli*. *Microbiology* **146**, 3157–3170.

Bailey, T. L. & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28–36.

Barnard, A., Wolfe, A. & Busby, S. (2004). Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr Opin Microbiol* **7**, 102–108.

Batut, J., Daveran-Mingot, M. L., David, M., Jacobs, J., Garnerone, A. M. & Kahn, D. (1989). *fixK*, a gene homologous with *fnr* and *crp* from *Escherichia coli*, regulates nitrogen fixation genes both positively and negatively in *Rhizobium meliloti*. *EMBO J* **8**, 1279–1286.

Bauer, C. E., Elsen, S. & Bird, T. H. (1999). Mechanisms for redox control of gene expression. *Annu Rev Microbiol* **53**, 495–523.

Bearson, S. M., Albrecht, J. A. & Gunsalus, R. P. (2002). Oxygen and nitrate-dependent regulation of *dmsABC* operon expression in *Escherichia coli*: sites for Fnr and NarL protein interactions. *BMC Microbiol* **2**, 13.

Christiansen, N. & Ahring, B. K. (1996). *Desulfitobacterium hafniense* sp. nov., an anaerobic, reductively dechlorinating bacterium. *Int J Syst Bacteriol* **46**, 442–448.

Christiansen, N., Ahring, B. K., Wohlfarth, G. & Diekert, G. (1998). Purification and characterization of the 3-chloro-4-hydroxyphenyla-cetate reductive dehalogenase of *Desulfitobacterium hafniense*. *FEBS Lett* **436**, 159–162.

Diaz, E. & Prieto, M. A. (2000). Bacterial promoters triggering biodegradation of aromatic pollutants. *Curr Opin Biotechnol* **11**, 467–475.

Egland, P. G. & Harwood, C. S. (2000). HbaR, a 4-hydroxybenzoate sensor and FNR–CRP superfamily member, regulates anaerobic 4-hydroxybenzoate degradation by *Rhodopseudomonas palustris*. *J Bacteriol* **182**, 100–106.

Gábor, K., Veríssimo, C. S., Cyran, B. C., Ter Horst, P., Meijer, N. P., Smidt, H., de Vos, W. M. & van der Oost, J. (2006). Characterisation of CprK1, a CRP/FNR-type transcriptional regulator of halorespiration from *Desulfitobacterium hafniense*. *J Bacteriol* **188**, 2604–2613.

Gostick, D. O., Griffin, H. G., Shearman, C. A., Scott, C., Green, J., Gasson, M. J. & Guest, J. R. (1999). Two operons that encode FNR-like proteins in *Lactococcus lactis*. *Mol Microbiol* **31**, 1523–1535.

Green, J., Scott, C. & Guest, J. R. (2001). Functional versatility in the CRP–FNR superfamily of transcription factors: FNR and FLP. *Adv Microb Physiol* **44**, 1–34.

Grimm, A. C. & Harwood, C. S. (1999). NahY, a catabolic plasmid-encoded receptor required for chemotaxis of *Pseudomonas putida* to the aromatic hydrocarbon naphthalene. *J Bacteriol* **181**, 3310–3316.

Harwood, C. S., Nichols, N. N., Kim, M. K., Ditty, J. L. & Parales, R. E. (1994). Identification of the *pcaRKF* gene cluster from *Pseudomonas putida*: involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate. *J Bacteriol* **176**, 6479–6488.

Hurles, M. (2004). Gene duplication: the genomic trade in spare parts. *PLoS Biol* **2**, E206.

Israelsen, H., Madsen, S. M., Vrang, A., Hansen, E. B. & Johansen, E. (1995). Cloning and partial characterization of regulated promoters from *Lactococcus lactis* Tn*917*-*lacZ* integrants with the new promoter probe vector, pAK80. *Appl Environ Microbiol* **61**, 2540–2547.

Janssen, D. B., Dinkla, I. J., Poelarends, G. J. & Terpstra, P. (2005). Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities. *Environ Microbiol* **7**, 1868–1882.

Joyce, M. G., Levy, C., Gábor, K., Pop, S. M., Biehl, B. D., Doukov, T. I., Ryter, J. M., Mazon, H., Smidt, H. & other authors (2006). CprK crystal structures reveal mechanism for transcriptional control of halorespiration. *J Biol Chem* **281**, 28318–28325.

Kolb, A., Busby, S., Buc, H., Garges, S. & Adhya, S. (1993). Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem* **62**, 749–795.

Korner, H., Sofia, H. J. & Zumft, W. G. (2003). Phylogeny of the bacterial superfamily of Crp–Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol Rev* **27**, 559–592.

Madsen, T. & Licht, D. (1992). Isolation and characterization of an anaerobic chlorophenol-transforming bacterium. *Appl Environ Microbiol* **58**, 2874–2878.

Mazon, H., Gábor, K., Leys, D., Heck, A. J., van der Oost, J. & van den Heuvel, R. H. (2007). Transcriptional activation by CprK1 is regulated by protein structural changes induced by effector binding and redox state. *J Biol Chem* **282**, 11281–11290.

McGinnis, S. & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **32**, W20–W25.

Mesa, S., Hennecke, H. & Fischer, H. M. (2006). A multitude of CRP/FNR-like transcription proteins in *Bradyrhizobium japonicum*. *Biochem Soc Trans* **34**, 156–159.

Nicholas, K. B., Nicholas, H. B., Jr & Deerfield, D. W., II (1997). GeneDoc: analysis and visualization of genetic variation. *EMBnet.news* **4** (2), 1–4.

Parales, R. E. & Harwood, C. S. (2002). Bacterial chemotaxis to pollutants and plant-derived aromatic molecules. *Curr Opin Microbiol* **5**, 266–273.

Pop, S. M., Kolarik, R. J. & Ragsdale, S. W. (2004). Regulation of anaerobic dehalorespiration by the transcriptional activator CprK. *J Biol Chem* **279**, 49910–49918.

Schell, M. A. (1985). Transcriptional control of the *nah* and *sal* hydrocarbon-degradation operons by the *nahR* gene product. *Gene* **36**, 301–309.

Schneider, T. D. & Stephens, R. M. (1990). Sequence Logos: a new way to display consensus sequences. *Nucleic Acids Res* **18**, 6097–6100.

Scott, C., Partridge, J. D., Stephenson, J. R. & Green, J. (2003). DNA target sequence and FNR-dependent gene expression. *FEBS Lett* **541**, 97–101.

Smidt, H. & de Vos, W. M. (2004). Anaerobic microbial dehalogenation. *Annu Rev Microbiol* **58**, 43–73.

Smidt, H., van Leest, M., van der Oost, J. & de Vos, W. M. (2000). Transcriptional regulation of the *cpr* gene cluster in *ortho*-chlorophenol-respiring *Desulfitobacterium dehalogenans*. *J Bacteriol* **182**, 5683–5691.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Utkin, I., Dalton, D. D. & Wiegel, J. (1995). Specificity of reductive dehalogenation of substituted *ortho*-chlorophenols by *Desulfitobacterium dehalogenans* JW/IU-DC1. *Appl Environ Microbiol* **61**, 346–351.

Van der Meer, J. R. & Sentchilo, V. (2003). Genomic islands and the evolution of catabolic pathways in bacteria. *Curr Opin Biotechnol* **14**, 248–254.

Van Eekert, M. H. & Schraa, G. (2001). The potential of anaerobic bacteria to degrade chlorinated compounds. *Water Sci Technol* **44**, 49–56.

Van Spanning, R. J., De Boer, A. P., Reijnders, W. N., Westerhoff, H. V., Stouthamer, A. H. & Van der Oost, J. (1997). FnrP and NNR of *Paracoccus denitrificans* are both members of the FNR family of transcriptional activators but have distinct roles in respiratory adaptation in response to oxygen limitation. *Mol Microbiol* **23**, 893–907.

Veldman, R., Reijnders, W. N. & Van Spanning, R. J. (2006). Specificity of FNR-type regulators in *Paracoccus denitrificans*. *Biochem Soc Trans* **34**, 94–96.

Villemur, R., Saucier, M., Gauthier, A. & Beaudet, R. (2002). Occurrence of several genes encoding putative reductive dehalogenases in *Desulfitobacterium hafniense/frappieri* and *Dehalococcoides ethenogenes*. *Can J Microbiol* **48**, 697–706.

Vollack, K. U., Hartig, E., Korner, H. & Zumft, W. G. (1999). Multiple transcription factors of the FNR family in denitrifying *Pseudomonas stutzeri*: characterization of four *fnr*-like genes, regulatory responses and cognate metabolic processes. *Mol Microbiol* **31**, 1681–1694.

Webre, D. J., Wolanin, P. M. & Stock, J. B. (2004). Modulated receptor interactions in bacterial transmembrane signaling. *Trends Cell Biol* **14**, 478–482.

Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev* **61**, 533–616.

Edited by: J. Green