Published online 7 July 2007

Nucleic Acids Research, 2007, Vol. 35, No. 14 ***4755–4766***
doi:10.1093/nar/gkm509

Probing key DNA contacts in AraR-mediated transcriptional repression of the *Bacillus subtilis* arabinose regulon

Irina Saraiva Franco[1], Luís Jaime Mota[1], Cláudio Manuel Soares[2] and Isabel de Sá-Nogueira[1,3,*]

[1]Laboratory of Microbial Genetics and [2]Laboratory of Protein Modeling, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa. Av. da República, Apt. 127, 2781-901 Oeiras and [3]Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Caparica, Portugal

Received May 11, 2007; Revised June 7, 2007; Accepted June 12, 2007

ABSTRACT

In the absence of arabinose, the AraR transcription factor represses the expression of genes involved in the utilization of arabinose, xylose and galactose in *Bacillus subtilis*. AraR exhibits a chimeric organization: the N-terminal DNA-binding region belongs to the GntR family and the C-terminal effector-binding domain is homologous to the GalR/LacI family. Here, the AraR–DNA-binding interactions were characterized *in vivo* and *in vitro*. The effect of residue substitutions in the AraR N-terminal domain and of base-pair exchanges into an AraR–DNA-binding operator site were examined by assaying for AraR-mediated regulatory activity *in vivo* and DNA-binding activity *in vitro*. The results showed that residues K4, R45 and Q61, located in or near the winged-helix DNA-binding motif, were the most critical amino acids required for AraR function. In addition, the analysis of the various mutations in an AraR palindromic operator sequence indicated that bases G_9, A_{11} and T_{16} are crucial for AraR binding. Moreover, an AraR mutant M34T was isolated that partially suppressed the effect of mutations in the regulatory *cis*-elements. Together, these findings extend the knowledge on the nature of AraR nucleoprotein complexes and provide insight into the mechanism that underlies the mode of action of AraR and its orthologues.

INTRODUCTION

The transcription factor AraR controls the utilization of carbohydrates in *Bacillus subtilis*. The control exerted by AraR is modulated by the presence of the effector molecule arabinose leading to induction of expression of at least 13 genes, comprising the arabinose (*ara*) regulon, which includes the *araR* gene (1–4). The products of these genes (*araABDLMNPQ-abfA*, *araE*, *abnA* and *xsa*) include extracellular and intracellular catabolic enzymes involved in the degradation of arabinose, galactose and xylose containing polysaccharides, uptake of these sugars into the cell and further catabolism of L-arabinose and arabinose oligomers (1–3,5).

A key property of AraR is its ability to bind specific DNA sequences in the absence of the inducer L-arabinose, as determined by DNAse I footprinting analysis (4,6,7). AraR recognizes and binds at least eight palindromic operator sequences, located in the five known arabinose-inducible promoters. Three of these promoters contain two *ara* boxes: the promoter of the *araABDLMNPQ-abfA operon* (boxes OR_{A1} and OR_{A2}), of *araE* (OR_{E1} and OR_{E2}) and of *xsa* (OR_{X1} and OR_{X2}). In the cases of the genes *araR* and *abnA*, a single *ara* box is present (OR_{R3} and OR_{B1}). AraR binding to the promoters displaying two *ara* boxes is cooperative, requiring in phase and properly spaced operators, and involves the formation of a small loop in the DNA. These two mechanistically diverse modes of action of AraR result in distinct levels of transcriptional regulation, as cooperative binding to two *ara* boxes results in a high level of repression while interaction with a single operator allows a more flexible control (4,6,7).

AraR is a 362 amino acid homodimeric protein that shows a chimeric organization, consisting of two functional domains with different phylogenetic origins (1,6,8): a small N-terminal DNA-binding domain (DBD) comprising a winged helix–turn–helix (HTH) motif belonging to the GntR family of transcriptional regulators (9)

*To whom correspondence should be addressed. Tel: +351 21 4469524; Fax: +351 21 4411277; Email: sanoguei@itqb.unl.pt
Present address:
Luís Jaime Mota. Imperial College London, Center for Molecular Microbiology and Infection, Armstrong Road, Flowers Building – 2nd floor, London SW7 2AZ, UK

and a larger C-terminal domain homologous to that of the GalR/LacI family of bacterial regulators and sugar-binding proteins (10). AraR typifies one of the six GntR-subfamilies of proteins (11,12). Currently, there are 54 members of this rapidly growing class of proteins, which can be found in prokaryotes [CDART database; (13)].

Previously, a model for AraR was derived using comparative modelling based on crystal structures of FadR (DBD) and PurR (COOH domain) from *Escherichia coli* (8). We have used random and site-directed mutagenesis to map the functional domains of AraR required for DNA binding, dimerization and effector binding. The arabinose-binding pocket is composed of polar and charged residues, whereas the dimerization interface has a hydrophobic nature. In both cases, the residues are distributed along the primary sequence of the C-terminal domain (8). Based on crystallographic studies of structurally and functionally related proteins, binding of the effector to the COOH region in AraR is predicted to elicit a conformational change in the N-terminal region, leading to inhibition of binding to operator sequences, and allowing transcription from the arabinose-responsive promoters. This allosteric signal involves a switching mechanism for communicating structural changes triggered in the sensor domain to the regulatory domain, decreasing the affinity of the latter for DNA.

Winged helix motifs are functionally and mechanistically versatile (14). They are primarily involved in DNA binding, but cases have been reported in which they participate in protein–protein interactions. Monomeric, homo- or heterodimeric protein–DNA complexes have been characterized and revealed quite distinct modes of binding to DNA, which can involve interactions between the recognition helix and the wing with the major and minor groove (15). Although the level of amino acid identity for the DBD of all members of the GntR superfamily is low (~25%) they share this conserved structural topology (11). Global analysis of the conservation of amino acid sequences in DNA-binding proteins concluded that residues interacting with the DNA backbone establish a set of core contacts that provide stability for homologous protein–DNA complexes, and consequently are well conserved across all protein families. On the other hand, residues that interact with DNA bases have more variable levels of conservation (16). Previous mutagenic studies showed that AraR residues in the N-terminal region were required for DNA binding because mutations in these residues abolished its regulatory function *in vivo* (8). However, the precise contribution of the mutated amino acids to DNA-binding activity was unclear.

To understand the specific properties of the interaction AraR-operator sequences, we substituted amino acids, in or near the HTH motif, which according to the model were predicted to contact DNA. We determined the effects of these substitutions on the ability of AraR to function *in vivo* and on the DNA-binding affinities *in vitro*. Conversely, mutational analysis of the AraR-binding sites was used to determine the base-specific requirements for transcriptional regulation *in vivo* and DNA binding *in vitro*. These experiments gave both expected and unexpected results, which together showed that specific AraR residues and operator bases are crucial to achieve a high level of regulatory activity, while others display variable contributions to DNA binding. In addition, an AraR mutant was isolated, which partially suppresses the loss of regulation observed in certain mutated DNA operators.

MATERIALS AND METHODS

Strains and growth conditions

Bacillus subtilis strains used in this work (Table 1) were grown in Luria–Bertani (LB) medium (17) or C minimal medium (18) and solid sugar-free agar (SFA) medium (LabM) or LB broth solidified with 1.6% agar. Chloramphenicol (5 μg ml^{-1}), kanamycin (10 μg ml^{-1}) and erythromycin (1 μg ml^{-1}) were added when appropriate. The Amy phenotype was tested by detection of starch hydrolysis on tryptose blood agar base medium (Difco) plates, containing 1% of potato starch, with a I_2–KI solution as described previously (3). *Escherichia coli* DH5α (Gibco BRL) or XL1-blue were used for routine molecular cloning work and *E. coli* BL21 DE3 pLysS (19) for overexpression of mutant AraR proteins. *Escherichia coli* strains were grown on LB medium, with ampicillin (100 μg ml^{-1}), chloramphenicol (20 μg ml^{-1}), kanamycin (30 μg ml^{-1}) and IPTG (isopropyl-β-D-thiogalactopyranoside) (1 mM) added as appropriate. The *B. subtilis* and *E. coli* cells were transformed as described previously (7).

DNA manipulations and sequencing

DNA manipulations were carried out as described previously (20). Restriction enzymes were purchased from MBI Fermentas, New England Biolabs or Roche, and used according to manufacturer's instructions. DNA was eluted from agarose gels using the GeneCleanII kit (Bio101) or the GFX DNA purification kit (GE Healthcare). PCRs were performed in a GeneAmp PCR system 2400 (Perkin-Elmer) and PCR products purified using QIAquick PCR purification kit (QIAgen). DNA was sequenced using an ABI PRIS BigDye terminator ready reaction cycle sequencing kit (Applied Biosystems).

Site-directed mutagenesis of *araR*

Amino acid substitutions in AraR were made by the QuikChange (Stratagene) site-directed method using as template plasmid pLS30 (8) and mutagenic oligonucleotides carrying the modified codon in the centre (listed in Supplementary Table 1). For R41A and H42A, a 486-bp region containing the mutations (BglII–MluI fragment) was subcloned in pLS30, originating plasmids pSC16 and pIF41, respectively. The linearized plasmids were used to transform *B. subtilis* strain IQB350, leading to the integration of the *araR* mutant alleles into the chromosome at the *amyE* locus via double recombination (Strains are listed in Table 1). Substitutions K4A, Y5F, E30A and Q61A were generated by site-directed

Table 1. *Bacillus subtilis* strains used in this work

Strain	Genotype	Relevant phenotype	Source
IQB101	*araAB′-lacZ erm*	Ara$^-$ LacZ$^-$	(1)
IQB350	*ΔaraR::km araAB′-lacZ erm*	Ara$^-$ LacZ$^+$	pLM8→IQB101[a,b]
IQB351	*ΔaraR::km araAB′-lacZ erm ΔamyE::araR cat*	LacZ$^-$	pLS24→IQB350
IQB352	*ΔaraR::km araAB′-lacZ erm ΔamyE::araR cat*	LacZ$^-$	pLS30→IQB350
IQB355	*ΔaraR::km araAB′-lacZ erm ΔamyE::araR F37S cat*	LacZ$^+$	pIF1→IQB350
IQB356	*ΔaraR::km araAB′-lacZ erm ΔamyE::araR Q61R cat*	LacZ$^+$	pIF2→IQB350
IQB357	*ΔaraR::km araAB′-lacZ erm ΔamyE::araR L33S cat*	LacZ$^+$	pIF3→IQB350
IQB358	*ΔaraR::km araAB′-lacZ erm ΔamyE::araR Δ13-65 cat*	LacZ$^+$	pIF8→IQB350
IQB505	*ΔaraR::km araAB′-lacZ erm ΔamyE::araR H42A cat*	LacZ$^+$	pIF41→IQB350
IQB513	*ΔaraR::km araAB′-lacZ erm ΔamyE::araR R41A cat*	LacZ$^-$	pSC16→IQB350
IQB563	*ΔaraR::km araAB′-lacZ erm ΔamyE::araR R45A cat*	LacZ$^+$	PCRmut R45A[c]→IQB350
IQB564	*ΔaraR::km araAB′-lacZ ermΔamyE::araR Q61A cat*	LacZ$^+$	PCRmut Q61A[c]→IQB350
IQB568	*ΔaraR::km araAB′-lacZ ermΔamyE::araR E30A cat*	LacZ$^+$	PCRmut E30A[c]→IQB350
IQB571	*ΔaraR::km araAB′-lacZ ermΔamyE::araR Y5F cat*	LacZ$^+$	PCRmut Y5F[c]→IQB350
IQB712	*ΔaraR::km araAB′-lacZ ermΔamyE::araR K4A cat*	LacZ$^+$	PCRmut K4A[c]→IQB350
IQB530	*ΔaraR::kmΔamyE::OR$_{A'}$-lacZ cat*	LacZ$^+$	pLM32→IQB215 (6)
IQB531	*ΔaraR::kmΔamyE::OR$_{A1\ (G9\ \rightarrow T)}$′ -lacZ cat*	LacZ$^+$	pLM56→IQB215 (6)
IQB532	*ΔaraR::kmΔamyE::OR$_{A1\ (A11\rightarrow C)}$′-lacZ cat*	LacZ$^+$	pLM64→IQB215
IQB533	*ΔaraR::kmΔamyE::OR$_{A1\ (T6\rightarrow G)}$′-lacZ cat*	LacZ$^+$	pLM65→IQB215
IQB534	*ΔaraR::kmΔamyE::OR$_{A1\ (T10\rightarrow G)}$′-lacZ cat*	LacZ$^+$	pLM66→IQB215
IQB535	*ΔaraR::kmΔamyE::OR$_{A1\ (T16\rightarrow G)}$′-lacZ cat*	LacZ$^+$	pLM67→IQB215
IQB536	*ΔaraR::kmΔamyE::OR$_{A1\ (A1\rightarrow C)}$′-lacZ cat*	LacZ$^+$	pLM68→IQB215
IQB537	*ΔaraR::kmΔamyE::OR$_{A1\ (G5\rightarrow T)}$′-lacZ cat*	LacZ$^+$	pLM69→IQB215
IQB538	*ΔaraR::kmΔamyE::OR$_{A2\ (G9\rightarrow T)}$′-lacZ cat*	LacZ$^+$	pLM78→IQB215
IQB257	*ΔaraR::kmΔamyE::OR$_{A1\ (C8\rightarrow A)}$′-lacZ cat*	LacZ$^\pm$	pLM70→IQB215
IQB572	*ΔaraR::kmΔamyE::OR$_{A}$′-lacZ cat thrC::araR erm*	LacZ$^\pm$	pIF76→IQB530
IQB573	*ΔaraR::kmΔamyE::OR$_{A1\ (G9\ \rightarrow T)}$′-lacZ cat thrC::araR erm*	LacZ$^+$	pIF76→IQB531
IQB574	*ΔaraR::kmΔamyE::OR$_{A1\ (A11\rightarrow C)}$′-lacZ cat thrC::araR erm*	LacZ$^+$	pIF76→IQB532
IQB575	*ΔaraR::kmΔamyE::OR$_{A1\ (T6\rightarrow G)}$′-lacZ cat thrC::araR erm*	LacZ$^+$	pIF76→IQB533
IQB576	*ΔaraR::kmΔamyE::OR$_{A1\ (T10\rightarrow G)}$′-lacZ cat thrC::araR erm*	LacZ$^\pm$	pIF76→IQB534
IQB598	*ΔaraR::kmΔamyE::OR$_{A1\ (C8\rightarrow A)}$′-lacZ cat thrC::araR erm*	LacZ$^\pm$	pIF76→IQB257
IQB599	*ΔaraR::kmΔamyE::OR$_{A1\ (T16\rightarrow G)}$′-lacZ cat thrC::araR erm*	LacZ$^+$	pIF76→IQB535
IQB700	*ΔaraR::kmΔamyE::OR$_{A1\ (A1\rightarrow C)}$′-lacZ cat thrC::araR erm*	LacZ$^+$	pIF76→IQB536
IQB701	*ΔaraR::kmΔamyE::OR$_{A1\ (G5\rightarrow T)}$′-lacZ cat thrC::araR erm*	LacZ$^+$	pIF76→IQB537
IQB702	*ΔaraR::kmΔamyE::OR$_{A2\ (G9\rightarrow T)}$′-lacZ cat thrC::araR erm*	LacZ$^+$	pIF76→IQB538
IQB583	*ΔaraR::kmΔamyE::OR$_{A}$′-lacZ cat thrC::araR M34T erm*		pIF85→IQB530
IQB708	*ΔaraR::kmΔamyE::OR$_{A1\ (G9\ \rightarrow T)}$′-lacZ cat thrC::araR M34T erm*		pIF85→IQB531
IQB709	*ΔaraR::kmΔamyE::OR$_{A1\ (A11\rightarrow C)}$′-lacZ cat thrC::araR M34T erm*		pIF85→IQB532
IQB582	*ΔaraR::kmΔamyE::OR$_{A1\ (T6\rightarrow G)}$′-lacZ cat thrC::araR M34T erm*		pIF85→IQB533
IQB710	*ΔaraR::kmΔamyE::OR$_{A1\ (T10\rightarrow G)}$′-lacZ cat thrC::araR M34T erm*		pIF85→IQB534
IQB704	*ΔaraR::kmΔamyE::OR$_{A1\ (C8\rightarrow A)}$′-lacZ cat thrC::araR M34T erm*		pIF85→IQB257
IQB703	*ΔaraR::kmΔamyE::OR$_{A1\ (T16\rightarrow G)}$′-lacZ cat thrC::araR M34T erm*		pIF85→IQB535
IQB705	*ΔaraR::kmΔamyE::OR$_{A1\ (A1\rightarrow C)}$′-lacZ cat thrC::araR M34T erm*		pIF85→IQB536
IQB706	*ΔaraR::kmΔamyE::OR$_{A1\ (G5\rightarrow T)}$′-lacZ cat thrC::araR M34T erm*		pIF85→IQB537
IQB707	*ΔaraR::kmΔamyE::OR$_{A2\ (G9\rightarrow T)}$′-lacZ cat thrC::araR M34T erm*		pIF85→IQB538

[a]The arrows indicate transformation and point from donor DNA to recipient strain.
[b]Transformation was always carried out with linearized DNA.
[c]Mutagenized pLS30 DNA was used as donor DNA (see Materials and Methods section).

mutagenesis of pLS30, the amplified products were ligated, linearized and used to transform *B. subtilis* (Table 1). Substitution R45L was obtained by chance when attempting to create mutation R45A (8) using an identical procedure. Presence of the mutations was verified by sequencing the *araR* allele in the resulting plasmids or strains.

Site-directed mutagenesis of operator regions

Plasmid pLM51 (6) a pBluescript II KS (+) derivative carrying the wild-type *araABDLMNPQ-abfA* promoter, was used as template for generating single-nucleotide substitutions in OR_{A1} or OR_{A2}, using the QuikChange (Stratagene) site-directed method and pairs of mutagenic oligonucleotides (listed in Supplementary Table 1). Resulting plasmids contained the following mutations in OR_{A1}: $A_1 \rightarrow C$ (pLM61), $G_5 \rightarrow T$ (pLM62), $T_6 \rightarrow G$ (pLM58), $C_8 \rightarrow A$ (pLM63), $G_9 \rightarrow T$ [pLM54; (6)], $T_{10} \rightarrow G$ (pLM59), $A_{11} \rightarrow C$ (pLM57), $T_{16} \rightarrow G$ (pLM60); or $G_9 \rightarrow T$ in OR_{A2} [pLM77; (7)]. The 204-bp BamHI–EcoRI DNA fragment from these plasmids, containing the mutagenized operator region, was then subcloned in the same sites of pSN32 (6). This procedure generated respectively pLM68, pLM69, pLM65, pLM70, pLM56 (6), pLM66, pLM64, pLM67 and pLM78 (7), which bear transcriptional fusions of the *araABDLMNPQ-abfA* promoter with single-point mutations to *lacZ*. After linearization, plasmids were used to transform *B. subtilis* IQB215, giving rise to strains where these fusions were integrated at the *amyE* locus (Table 1). To analyse the

repression exerted by AraR on the *ara* boxes, integration of the *araR* allele at the *thrC* locus was accomplished by transformation with pIF76.

Isolation of AraR suppressor mutants

The insertion of a 1446-bp *Eco*RI–*Bam*HI fragment from pLS30 (8), containing the *araR* allele, into pDG1664 (EcoRI–BamHI) (21) yielded plasmid pIF76. This plasmid was used as DNA template in random PCR mutagenesis, according to the method described by Leung *et al*. (22). Random PCR mutagenesis with oligos ARA6 and ARA73 amplified a 650-bp 5′-end region of the *araR* allele. After digestion with BamHI–Eco47III, the fragment was subcloned in the same plasmid leading to the replacement of the equivalent region of the wild-type *araR* allele, from sites −227 to +251 relative to the transcription start site (containing the promoter and the first 75 codons of the *araR* gene). The recombinant plasmids were transformed into *E. coli* DH5α yielding a library of *araR* mutations (contained in ∼2200 transformants). A plasmid pool was used in separate experiments to transform *B. subtilis* strains with a Δ*araR* Δ*amyE::OR$_A$-lacZ* background, in which the OR_A operator sequence carried the mutations described above (Table 1), leading to integration of AraR mutants at the *thrC* locus via double crossing-over. The constitutive expression of *lacZ* due to the presence of the mutated *ara* boxes leads to a Lac^+ phenotype in the receptor strains, reflected by a blue colour in SFA medium with X-gal. To isolate mutant *araR* alleles suppressing the deleterious effect of the operator mutations, we screened for colonies displaying a weaker Lac phenotype (white/light blue phenotype in the same medium). Chromosomal DNA from these colonies was used as template to amplify the mutagenized region of the *araR* allele, which was subsequently cloned back into pIF76 as described above and sequenced. The resulting plasmid, pIF85, bears a mutation leading to a single amino acid substitution, M34T.

β-Galactosidase assays

Bacillus subtilis strains were grown in C minimal medium supplemented with 1% (w/v) casein hydrolysate in the presence and in the absence of L-arabinose 0.4% (w/v) as previously reported (1). Samples of cell culture were collected and analysed 2 h after the addition of L-arabinose. The ratio of β-galactosidase activity, determined as described (17) from cultures grown for 2 h in the presence and absence of inducer was taken as a measure of AraR repression in the analysed strains (Repression Index).

Imunoblotting of cell extracts

Bacillus subtilis strains were grown as for β-galactosidase assays. Preparation of cell extracts and immunoblotting were performed as described (8). Blots were developed with anti-AraR-MBP2* serum (6) using the ECL detection system (Amersham Biosciences). Protein concentration was determined using a Bio-Rad kit.

Construction of plasmids for overexpression of AraR mutants and protein purification

Fusion of the C-terminus of AraR variants to six histidines in the plasmid pET30a(+) (Novagen) was engineered, placing the genes under the control of a T7 promoter. The construction of plasmid pLS16, carrying the wild-type allele, was described previously (7). Construction of plasmids carrying the AraR substitutions F37S, Q61R and L33S was accomplished in a similar manner. Briefly, the alleles containing these mutations were amplified with oligos ARA50 and ARA51 from the pLS30-derivatives pIF1, pIF2 and pIF3 obtained previously (8). The 1112-bp PCR product was separately digested with AvaI–NdeI and AvaI–HindIII, the resulting 282-bp and a 805-bp fragments were inserted in pET30a(+) restricted with NdeI and HindIII, yielding plasmids pIF5, pIF6 and pIF7, respectively. To introduce mutations S53P, H42A or M34T, regions BglII–KpnI were obtained from plasmids pIF17, pIF41 and pIF85 described above, and used to substitute the same region in pIF7 generating pIF111, pIF123 and pIF121. For K4A, Y5F, E30A, R45A and Q61A *B. subtilis* chromosomal DNA from strains IQB712, IQB571, IQB568, IQB563 and IQB564 was used (Table 1). PCR products were digested with appropriate enzymes (NdeI–KpnI, BglII–KpnI or BglII–HindIII) and used to substitute the corresponding region in pIF7. These procedures yielded pIF124, pIF112, pIF78, pIF74 and pIF75, respectively. The presence of the mutations was verified by sequencing the *araR* alleles. For the purification of these AraR-his6 variants, *E. coli* BL21 (λDE3) pLysS (19) cells transformed with the corresponding pET30 derivatives were grown at 37°C to an optical density at 600 nm of 0.6 in 1 l of LB medium, and then expression of the fusion proteins was induced by addition of IPTG to 1 mM. Incubation in the same conditions continued for additional 2 h. All subsequent steps were carried out similarly to the method described previously (7).

Electrophoretic mobility shift assays (EMSAs)

A DNA fragment carrying the OR_{A1}–OR_{A2} region was amplified from pLM51 using primers ARA262 and ARA263 (Supplementary Table 1). After purification, the 126-bp PCR product was labelled with T4 Polynucleotide Kinase (MBI Fermentas) and [γ-32P] dATP, followed by extraction with phenol/chloroform and precipitation with ethanol. Binding reactions contained 12 mM HEPES-KOH pH 7.6, 10 mM $MgCl_2$, 0.5% [w/v] BSA, 1 mM DTE, 10% Glycerol (v/v), 200 mM NaCl, 4 mM Na_2HPO4, 4 mM NaH_2PO4, 0.4 mM EDTA, a 200-fold molar excess of competitor DNA (polydIdC), 1 nM of labelled DNA and increasing concentrations of wild-type or mutant AraR proteins. After incubation for 30 min at room temperature, the mixture was loaded onto a pre-run 8% polyacrylamide gel in 25 mM Tris 200 mM Glycine (pH 8.9) and run at 100 V for ∼1 h. Gels were dried under vacuum and exposed to a Phosphorimager screen before analysis with a Molecular Dynamics Storm 860 Imager and ImageQuant version 5.0. To determine the dissociation constants, protein concentrations were used

according to previous *in vitro* results (6) and K_d values were obtained using the GraphPad Prism software. For competition DNA-binding experiments, various amounts of cold double-stranded oligonucleotides containing single mutations in the operator sequences (Supplementary Table 1) were added to the reaction in the presence of 40 nM AraR. As controls, we used oligonucleotides carrying the wild-type operator (ARA288 and ARA289) or a non-specific DNA sequence with the same length (ARA244 and ARA245). The following procedures were made as described above. The percentage inhibition in the presence of competitor DNA was determined similarly to the method described by Bera *et al.* (23). The radioactivity of bound DNA was quantified in the control without competitor, and in samples containing 500-fold molar excess of the distinct competitors. Inhibition (%) = $100 \times [(\text{bound})_{\text{control}} - (\text{bound})_{\text{sample}}/(\text{bound})_{\text{control}}]$.

RESULTS AND DISCUSSION

In vivo effect of amino acid substitutions in the DBD of AraR

We have previously established that AraR interaction with DNA is achieved via a 70 amino acid N-terminal domain. These results were obtained by random and site-directed mutagenesis based on a 3D model of the DBD derived from the crystal structure of the *E. coli* regulator FadR (8). However, many of these mutations resulted in changes that could alter the protein structure or interfere with DNA binding. In order to conduct a more clarifying characterization of the role of specific AraR residues on its DNA-binding activity, we made several amino acid substitutions in or near the HTH motif. Residues were chosen based on the 3D model and/or primary sequence alignment of AraR-like proteins (8). The majority of the substituted amino acids was predicted to contact directly the bases of the DNA and consequently, would account for the specificity of the interaction with operator sequences. Positions K4, E30, R41, H42 and Q61 were exchanged to alanine and Y5 to phenylalanine (Figure 1A and B). The substitutions were designed to minimize local structure disruption and probe loss of contact with the DNA. In addition, in this work we analysed a mutation R45L generated by chance during the construction of mutant R45A, which was characterized *in vivo* in a previous work (8). Plasmid pLS30, carrying an *araR* wild-type allele, and its derivatives harbouring the mutated *araR* alleles, obtained after site-directed mutagenesis, were integrated as single copy at the *amyE* locus of *B. subtilis* receptor strain IQB350. This strain bears an *araR* null mutation (Δ*araR*) and a transcriptional *araAB′–lacZ* fusion (see Materials and Methods section) and therefore expresses constitutively β-galactosidase. In the resulting strains (Table 1), the levels of β-galactosidase expressed from the *ara* operon-*lacZ* fusion reflect the *in vivo* regulatory activity of the AraR variant encoded by the allele integrated at the *amyE* locus.

The effect of each substitution was analysed by determining the levels of accumulated β-galactosidase in strains grown under inducing (presence of arabinose) and non-inducing conditions (absence of arabinose). The results are summarized in Figure 2A. The regulatory activity was quantified as the repression index (Figure 2A). The receptor strain IQB350 (Δ*araR araAB′-lacZ*) and IQB352, a derivative carrying the *araR* wild-type allele at the *amyE* locus, were used as controls, yielding repression index values of 1 and 99, respectively, which correspond to the absence and maximal regulation exerted by the protein. Mutation R41A had no effect on the regulatory activity when compared to the wild type. Therefore, although amino acid R41 is conserved among all members of the GntR family (Figure 1A) and establishes interactions with bases in the major groove according to FadR-DNA data [R45 in FadR; (24)], it is dispensable for AraR binding *in vivo*. Variant H42A displayed a minimal decrease on repression activity

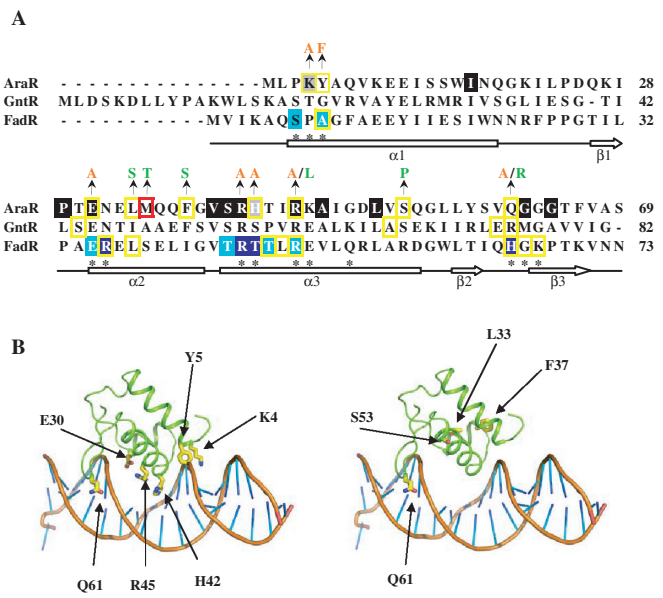


Figure 1. The DNA-binding domain of AraR and localization of mutations. (**A**) Sequence alignment of the N-terminal region of AraR, FadR and GntR. Residues that are conserved in the entire family GntR are shaded in black and residues characteristic of AraR homologous proteins in grey (8). Positions of mutations leading to a constitutive phenotype are boxed in yellow (8,26,29). A substitution yielding a suppressor phenotype is boxed in red. The introduced residues in AraR are shown above the sequence, coloured in orange when obtained through site-directed mutagenesis and green through random mutagenesis (8). In FadR are coloured in light blue residues contacting the DNA backbone and in dark blue the ones contacting the DNA bases, according to crystallographic data (28); an asterisk below the sequence indicates amino acids within contact distance of DNA (24). The secondary structure (arrows representing beta-strands and bars alpha-helices) of FadR (amino acid residues 1–73) is shown below the alignment according to van Aalten *et al.* (24). The microorganisms of source and accession numbers are: AraR from *B. subtilis* (P96711); GntR from *B. subtilis* (P10585); FadR from *E. coli* (P09371). (**B**) Structure of the modelled N-terminal domain of AraR (depicted in green ribbons; only one monomer is represented for clarity; see (8) for details) together with the DNA segment (depicted in orange ribbons) crystallized with FadR, highlighting the site-directed (left) and random (right) mutations displaying a constitutive phenotype. Site-directed mutations: K4→A, Y5→F, E30→A, H42→A, R45→A and Q61→A. Random mutations: L33→S, F37→S, S53→P and Q61→R.

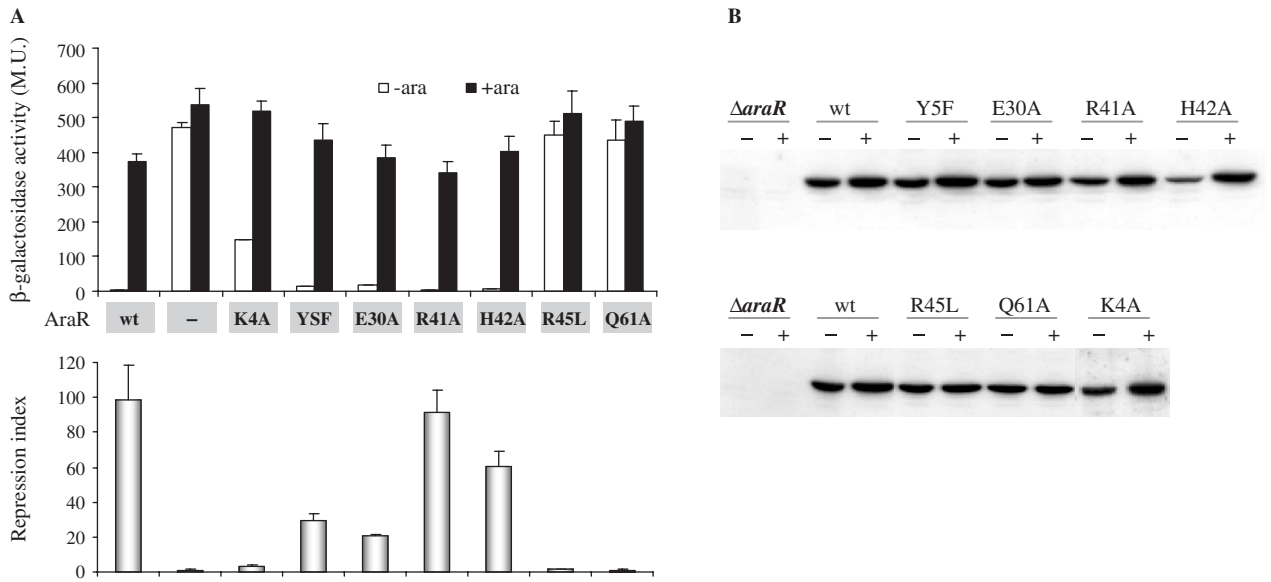


Figure 2. *In vivo* characterization of AraR mutants. (**A**) Regulatory activity of mutant AraR proteins. β-Galactosidase activities of *B. subtilis* strains carrying an *araAB′–lacZ* fusion and an *araR* allele integrated at the *amyE* locus were determined after growth in the absence or presence of inducer (upper panel, in white and black bars, respectively). Amino acid substitutions (obtained by site-directed mutagenesis) are indicated for the mutated position and amino acid substituted using the standard one-letter designation. The repression index (lower panel), calculated as the ratio between values obtained in the presence and in the absence of inducer, reflects the regulation exerted by each protein variant. Values are the average of three independent experiments, each assayed in duplicate. Error bars represent the SD. M.U.—Miller Units. Results obtained with strain harbouring wild-type AraR and null-mutant were previously reported (8). (**B**) AraR accumulation in the cell by western immunoblot analysis. Equal amounts of the soluble fractions of cell extracts obtained from the same *B. subtilis* strains grown in the absence (−) or presence (+) of inducer, were prepared as described in the Materials and Methods section.

(1.6-fold). Moderate effects were observed with Y5F and E30A, 3.3- and 4.8-fold, respectively. The more drastic effects were seen with K4A, a 30-fold decrease in repression, and with R45L and Q61A the regulatory activity was completely abolished. The lack of regulation of R45L is identical to that observed with R45A (8). Together, the results suggest that these three residues play the most important roles in DNA binding of the ones analysed here.

Because the observed decrease in repression could be the result of deficient *in vivo* accumulation of the mutant proteins, as a consequence of lower stability and proteolysis, we measured the abundance of each AraR variant. The strains were grown as for the β-galactosidase assays and the level of AraR estimated by western immunobloting using equivalent amounts of their soluble cell extracts (Figure 2B). The cellular level of all mutant proteins was comparable to that seen with wild-type AraR, ruling out the possibility of deregulation originated by degradation of the repressor.

Analysis of the wild-type AraR and mutant proteins DNA-binding affinity *in vitro*

The apparent affinity constants (K_d) of AraR mutants for operator sequences were determined by EMSAs using a ^{32}P-labelled 126-bp DNA fragment, which carried both operators of the metabolic operon, OR_{A1} and OR_{A2} (depicted in Figure 5). Binding of AraR to this DNA fragment was specific, as the presence of the inducer arabinose but not xylose prevented the formation of the protein–DNA complex (Figure 3A). Titration of 1 nM of DNA with increasing concentrations of wild-type repressor (Figure 3B) allowed the determination of an apparent K_d 3.9×10^{-8} M, which is defined as the amount of protein necessary to shift 50% of the labelled probe (25). This value is comparable to that previously calculated for each individual box using DNase I quantitative footprinting experiments, 3.4×10^{-8} M and 4.7×10^{-8} M for OR_{A1} and OR_{A2}, respectively (6).

All the mutant proteins that displayed an effect *in vivo* were overexpressed in *E. coli* and purified to homogeneity (see Materials and Methods section). Binding to DNA was assayed by EMSA and their respective apparent affinity constants determined (Figure 3C). Variant H42A, which showed the minimal loss of repression *in vivo* (2-fold) bound DNA with an apparent K_d of 4.1×10^{-8} M, similar to the wild-type protein. The most severe effects were displayed by Q61A ($K_d \sim 5.5 \times 10^{-7}$ M), K4A and R45A, both showing an apparent $K_d > 1.5 \times 10^{-6}$ M. These three mutants were also unable to perform a regulatory activity *in vivo*. Residues K4 and H42 are completely conserved among AraRs (8) and in a contact distance of the DNA according to the model (Figure 1A and B), however, mutation of these residues had different outcomes. Since these two residues are not conserved among the members of the entire GntR family they may contribute in very different extents to the DNA-binding specificity of AraR-like proteins.

R45 is a conserved amino acid in the GntR family members. In the regulator FadR from *E. coli*, the

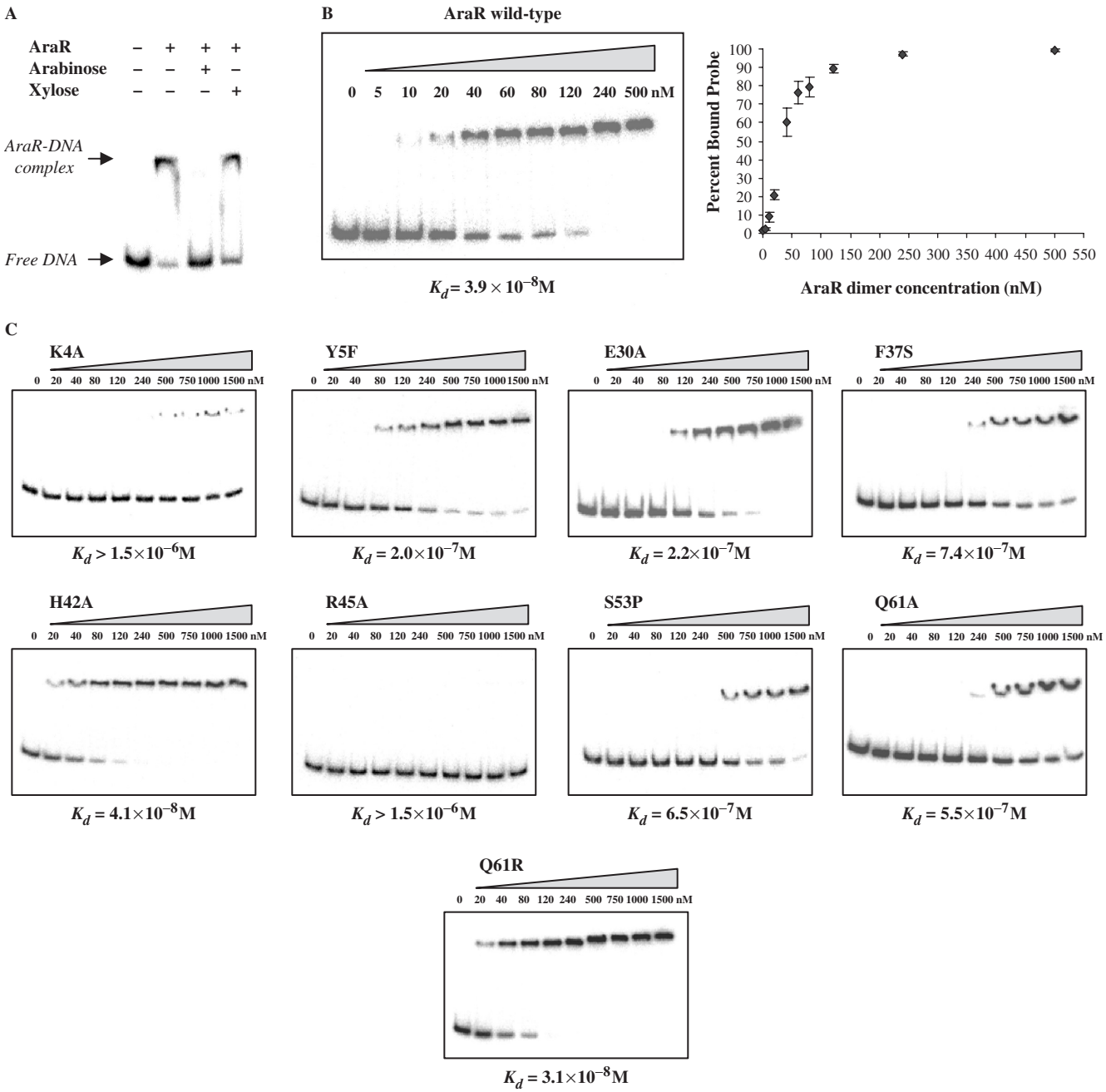


Figure 3. Binding of AraR to *araABDLMNPQ-abfA* promoter (operators OR_{A1}-OR_{A2}) in EMSA. AraR was incubated with the 5′-end-labelled probe (1–2 nM) and the protein–DNA complexes resolved by electrophoresis on native 8% polyacrylamide gels. Protein concentrations were calculated considering a pure dimeric protein. (**A**) Specificity controls for AraR binding. AraR (60 nM) was incubated with the DNA probe in the presence of L-arabinose or D-xylose (15 mM). (**B**) The indicated amounts of wild-type AraR were used in the binding reactions (left). Densitometric quantification of the bands corresponding to free DNA and protein–DNA complex allowed the determination of the affinity constant (right) (see Materials and Methods section). The values shown represent the average and standard deviation of at least three independent assays, with an intrinsic error <30%. (**C**) Effect of AraR substitutions on binding to DNA probe. The indicated concentrations of the mutant proteins were used in EMSA and determination of K_d was made as described above.

substitution of the corresponding residue (R49) has also a drastic effect *in vivo* (26). Moreover, the crystal structure of the FadR–DNA complex (27,28) shows that R49 locates in the recognition helix of the winged HTH and interacts with a phosphate group, not specifically a base. According to the predictions of the tertiary structure of AraR, R45 is also located in the recognition helix (Figure 1B), which is generally more responsible for the interaction with DNA, in particular the positively charged residues. Q61 belongs to the predicted wing of the DNA-binding motif (Figure 1A and B). The corresponding residue in both FadR and in GntR from *B. subtilis* is also positively charged (Figure 1A) and substitutions led to loss of DNA-binding ability (26,29). In FadR, H65

is part of the wing and makes specific contacts with an adenine (24,28).

Intermediate decreases of DNA binding were observed with Y5F and E30A, with K_d 2.0×10^{-7} M and 2.3×10^{-7} M, respectively, similarly to that seen *in vivo*. Therefore, these exchanges led to a comparable effect both *in vivo* and *in vitro* (Figures 2 and 3). Moreover, the nature of the mutation Y5F revealed the importance of the OH group in the interaction with DNA. Both residues are conserved in the GntR-family proteins, and the corresponding residues in FadR, A9 and E34, were shown to contact the DNA backbone (24,28). The latter also contacts nearby amino acids, contributing presumably to the stabilization of residues that interact specifically with the DNA bases.

Additionally, four other AraR mutants, L33S, F37S, S53P and Q61R, obtained by random mutagenesis and characterized *in vivo* in a previous work [(8); Figure 1B), were studied by EMSA. Both substitutions F37S and S53P had led to derepression *in vivo*, of ~24- and 2-fold, and in mutant L33S the regulatory activity was almost completely abolished (8). These values could be explained by the observed instability of the proteins (8). However, purified mutants F37S and S53P showed decreased DNA-binding affinities, K_d 7.4×10^{-7} M and 6.5×10^{-7} M, respectively, that may also contribute to the deregulation *in vivo*. These results could be explained by the nature and localization of these substitutions, which suggest implications in the folding of the DBD. Overexpression and purification of L33S yielded only small amounts of protein. Nevertheless, at the maximal concentration that we could use in EMSA assays, 50 nM of the mutant, no DNA binding was observed (data not shown).

Interestingly, while the Q61A substitution completely abolished regulation *in vivo* and DNA binding *in vitro*, the change to arginine in the same position showed only a 1.6-fold decrease of regulatory activity *in vivo* (8), and the affinity to the DNA probe *in vitro*, K_d 3.1×10^{-8} M, was even slightly higher than that displayed by wild-type AraR. Noteworthy, in GntR the inverse of AraR mutation Q61R (i.e. GntR R75Q) leads to a significant loss of regulation *in vivo* (29). Based on these observations, we may speculate that the rise of positive charge as a result of AraR substitution Q61R increased the overall (non-specific) affinity for DNA, leading *in vivo* to a titration of the protein.

In summary, K4, R45 and Q61, were the most critical AraR residues in achieving specific DNA binding, and Y5 and E30 also play an important role, and overall there was a good correlation between the effects of the mutations in the binding affinities to the *ara* operon promoter *in vitro* and in the regulatory activity *in vivo*.

Effect of base-pairs mutations in the operator sequences on transcriptional regulation by AraR

AraR recognizes and binds at least eight palindromic operator sequences, located in the five known arabinose-inducible promoters. Three of these promoters contain two *ara* boxes: the promoter of the *ara* metabolic operon (boxes OR_{A1} and OR_{A2}), of *araE* (OR_{E1} and OR_{E2}) and of *xsa* (OR_{X1} and OR_{X2}). In the cases of the genes *araR* and *abnA*, a single *ara* box is present (OR_{R3} and OR_{B1}). AraR binding to the promoters displaying two *ara* boxes is cooperative and involves the formation of a small loop in the DNA. In fact, for full *in vivo* repression, communication between repressor molecules bound to two properly spaced operators is required, as shown by the analysis of mutations designed to prevent cooperative binding of AraR (6,7). An alignment of the eight *ara* boxes, identified by DNase I footprinting and/or mutagenesis, showed the 16-bp consensus sequence 5′-ATTTGTACGTACAAAT-3′ and highlighted the conserved nucleotides at each position (Figure 4A). This operator consensus presents the typical signature for *cis*-acting elements recognized by GntR family members 5′-(N)x-GT-N(0-15)-AC-(N)x-3′ (11).

In a previous work, we showed that G_9 is important for AraR binding because the substitution $G_9 \rightarrow T$ in both boxes OR_{A1} and OR_{A2} caused defect in the regulatory activity of AraR *in vivo* and prevented cooperative binding (6). To further investigate which nucleotides within the consensus sequence were necessary for protein binding, single-nucleotide exchanges were made in OR_{A1} at the promoter of the *ara* operon. The most conserved bases were substituted and mutations were designed to introduce transversions from AT to CG and CG to AT: $A_1 \rightarrow C$, $G_5 \rightarrow T$, $T_6 \rightarrow G$, $C_8 \rightarrow A$, $T_{10} \rightarrow G$, $A_{11} \rightarrow C$ and $T_{16} \rightarrow G$. The mutated promoters transcriptional fused to the *lacZ* gene were independently integrated at the *amyE* locus of the *B. subtilis* receptor strain (see Materials and Methods section and Table 1). Strain IQB572, bearing a transcriptional fusion to the wild-type operator (*ΔaraR ΔamyE::*$OR_{A(wt)}$*-lacZ thrC::araR)* was used as a control to assay the repression exerted by AraR. The levels of accumulated β-galactosidase activity measured in all strains are shown in Figure 4B. Mutations having the most drastic effect on AraR binding were $G_9 \rightarrow T$, both in OR_{A1} and OR_{A2} [as previously determined; (7)], $A_{11} \rightarrow C$ and $T_{16} \rightarrow G$, leading to a decrease in the regulatory activity >9-fold compared to the control. A moderate effect of deregulation, varying from 2.4- to 4.4-fold, was observed for $A_1 \rightarrow C$, $G_5 \rightarrow T$ and $T_6 \rightarrow G$, and substitution $T_{10} \rightarrow G$ had no effect *in vivo*.

Surprisingly, $C_8 \rightarrow A$ abolished *lacZ* expression both in inducing and non-inducing conditions. One possibility to explain this result would be an increase in the affinity of AraR for the mutated operator leading to a tight binding of the repressor (even in the presence of arabinose), thus preventing transcription by RNA polymerase. To test this hypothesis, we investigated the Lac phenotype in *araR*-null mutants (*ΔaraR*) bearing the transcriptional fusion $OR_{A1(C8 \rightarrow A)}$–*lacZ* and $OR_{A(wt)}$–*lacZ*, strains IQB257 and IQB530 (Table 1), respectively. The results obtained in solid medium with X-gal and arabinose indicated that the lack of *lacZ* expression in mutant $C_8 \rightarrow A$ (Position +4, relative to the transcriptional start site) was independent of the presence of AraR (data not shown) suggesting that the mutation may affect transcription initiation by RNA polymerase.

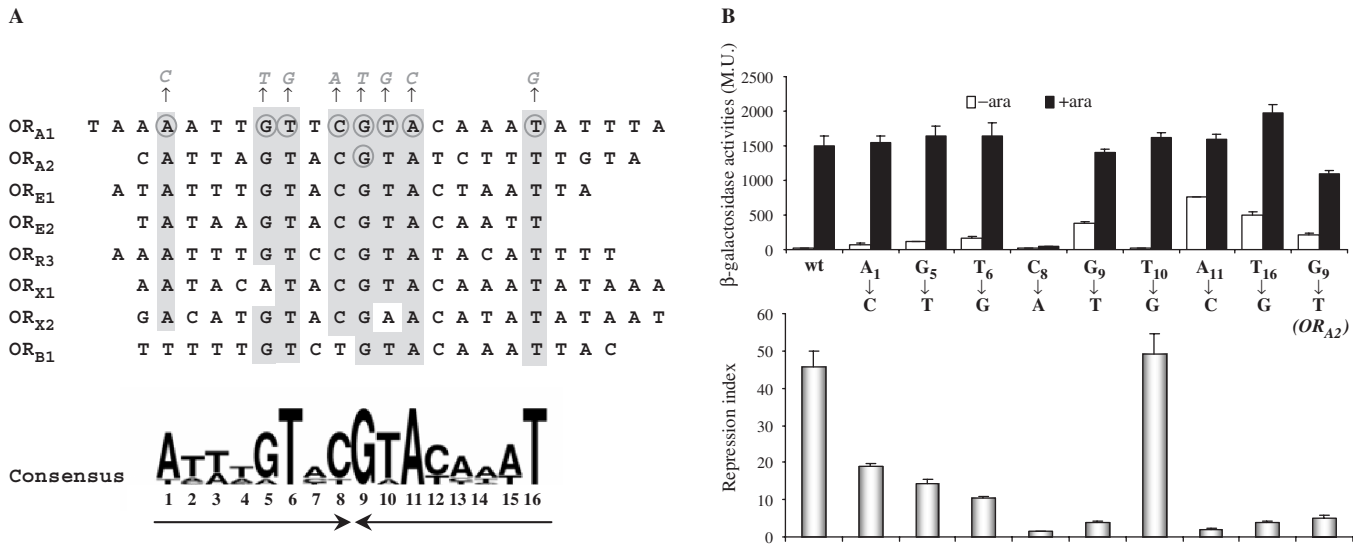


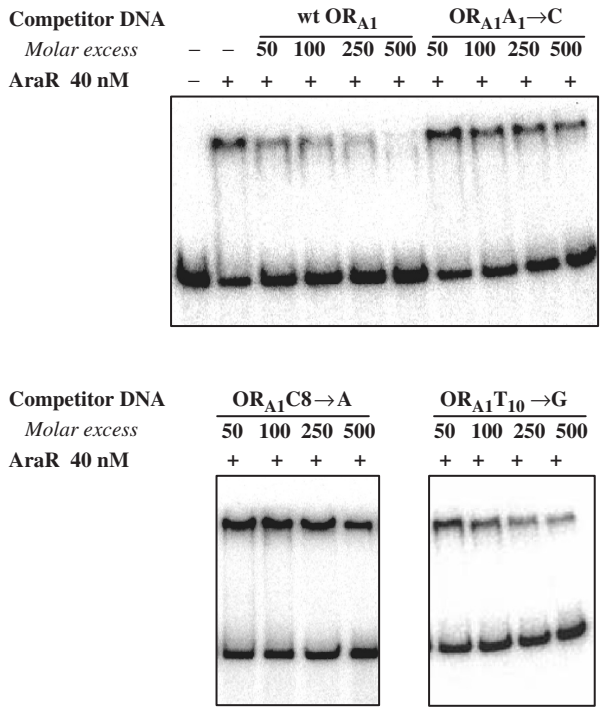
Figure 4. Ara operator sequences. (**A**) Alignment of the eight AraR boxes and a picture representing the conservation of bases at each position in the inferred consensus 16-bp palindromic operator sequence [generated by WebLogo 2.8.2 software (33)]. *Ara* boxes are located in the promoter of the *araABDLMNPQ-abfA* metabolic operon (OR_{A1} and OR_{A2}), *araE* (OR_{E1} and OR_{E2}), *araR* (OR_{R3}), *xsa* (OR_{X1} and OR_{X2}) and *abnA* (OR_{B1}). The conserved nucleotides in at least seven boxes are shaded. Bases substituted in OR_{A1} or OR_{A2} by site-directed mutagenesis used for *in vivo* analysis are indicated, and the new base is shown above. (**B**) Effect of *ara* box mutations on *in vivo* regulation by AraR. β-Galactosidase activities were determined in *B. subtilis* strains carrying ORA–*lacZ* fusions integrated at the *amyE* locus and wild-type *araR* at the *thrC* locus grown in the absence or presence of arabinose (upper panel, in white and black bars, respectively). The repression index (lower panel) reflects the regulation exerted on each *ara* box (Figure 2). Nucleotide substitutions (obtained by site-directed mutagenesis) are indicated for the mutated position in the *ara* consensus box (Figure 4A). Values are the average of three independent experiments, each assayed in duplicate. Error bars represent the SD. M.U.—Miller Units.

Analysis of operator mutations on AraR–DNA affinity *in vitro*

The effect of the operator mutations was also analysed *in vitro* by EMSA competition assays. The experiments were performed in the presence of 1 nM of the OR_{A1}–OR_{A2} DNA probe described above, 40 nM of AraR, and increasing concentrations of a double-stranded 38-bp competitor oligonucleotide (50–500 nM) containing the wild-type or the described mutations in OR_{A1} or OR_{A2}. In addition, oligonucleotides carrying all possible substitutions at the highly conserved base pairs, G_9, A_{10} and T_{16} were also used. We compared the ability of these cold DNAs to titrate binding of AraR to the labelled probe, reflected in the decrease of the intensity of the protein–DNA complex band. The wild-type OR_{A1} box was able to compete for AraR binding in a concentration-dependent manner, with a 79% loss of band shift at 500-fold excess competitor DNA (Figure 5A). In contrast, a non-specific oligonucleotide (equivalent in length; Supplementary Table 1) used as control disrupted only 18% of the binding (data not shown). Inhibition of binding in the presence of 500 nM of the different competitors was quantified and the results are summarized in Figure 5B. The DNA containing mutation $T_{10}\rightarrow G$ in OR_{A1} competed in levels similar to that obtained for the wild-type box (68 and 79% inhibition, respectively). In contrast, AraR was unable to bind the boxes with single base-pair substitutions in G_9, either to T (the mutation tested *in vivo*, previously), to A and C (inhibition values between 21 and 16%). A notorious decrease in binding to $A_{11}\rightarrow C$ was also observed, which was more pronounced when A was exchanged for G or T. However, oligonucleotides containing a mutation at T_{16} (either to G, A or C) were still able to partially compete for the repressor. The three mutations, $A_1\rightarrow C$, $G_5\rightarrow T$ and $T_6\rightarrow G$, leading to a partial de-repression *in vivo* also showed an intermediate effect. Taken together, the results indicate a good correlation between the *in vivo* and *in vitro*, but the exchanges at T_{16} and the mutation $A_{11}\rightarrow C$, comparatively to the regulatory activity *in vivo* were expected to bind to AraR less tightly. This could be due to the more reduced sensitivity of the competition assays. Interestingly, the mutation $C_8\rightarrow A$, that affected transcription even in the presence of inducer, did not compete (23%) for AraR, indicating that the mutation has an effect on AraR binding that could not be tested *in vivo*.

In previous work, a search for AraR operator sequences in the *B. subtilis* genome (6,30) identified a putative binding sequence in the open reading frame *ydjK* [identified as a myo-inositol transporter, *iolT*; (31)]. The sequence 5′-TTTTTACGTACAATT-3′ [+27 relative to the translation start site; (31)] displayed only two deviations (underlined) from the consensus sequence: A_1 and G_5. Construction and analysis of transcriptional fusions of the promoter region and 5′-end of *ydjK* to *lacZ* showed that expression is not AraR dependent, thus the potential operator is not functional *in vivo* (Inácio,J.M. and I.S.-N., unpublished data). To determine the ability of AraR to bind this sequence *in vitro*, competition assays were performed as described above but no competition was detected (Figure 5B). Since T at Position 1 is present in functionally active AraR boxes (Figure 4A) these observations suggest, in accordance to the mutagenic

A



B

Competitor DNA	Inhibition of AraR binding (%)
OR_{A1} **wt**	79 ± 15
Non-specific DNA	18 ± 3
OR_{A1} **A_1→C**	46 ± 13
OR_{A1} **G_5→T**	35 ± 5
OR_{A1} **T_6→G**	36 ± 11
OR_{A1} **C_8→A**	23 ± 3
OR_{A1} **G_9→T**	21 ± 1
OR_{A1} **G_9→A**	16 ± 2
OR_{A1} **G_9→C**	16 ± 1
OR_{A1} **T_{10}→G**	68 ± 12
OR_{A1} **A_{11}→C**	39 ± 12
OR_{A1} **A_{11}→G**	21 ± 3
OR_{A1} **A_{11}→T**	28 ± 9
OR_{A1} **T_{16}→G**	48 ± 0
OR_{A1} **T_{16}→A**	40 ± 12
OR_{A1} **T_{16}→C**	48 ± 2
OR_{A2} **G_9→T**	35 ± 7
ydjK	25 ± 7

Figure 5. *In vitro* analysis of AraR binding to mutated ara boxes. (**A**) Competition EMSA experiments using double-stranded oligonucleotides containing mutated *ara* boxes. AraR (40 nM) was incubated with the ^{32}P-labelled OR_{A1}–OR_{A2} region (1 nM) in the presence or absence of the indicated molar excess of 38-bp ds oligonucleotide competitors with wild type or the mutated *ara* boxes shown above. Protein–DNA complexes were resolved by electrophoresis in 8% polyacrylamide gels. Representative results are shown. (**B**) Quantification of the inhibition of AraR binding to wild-type operator sequence in the presence of competitor DNA. The values represent the percentage of inhibition of AraR binding to the labelled DNA probe observed in the presence of 500-fold molar excess of competitor DNA. For quantification, the intensity of the bands corresponding to protein–DNA complexes in EMSA, obtained in the presence or absence of competitor, were quantified in a densitometer. The percentage of inhibition was calculated as described in the Materials and Methods section. The results are the average and SD of at least three independent experiments, with a maximal associated error of 30%.

analysis described above, that G at Position 5 plays an important role in AraR binding.

In conclusion, we found that bases in both arms of the palindrome of the AraR boxes are involved in AraR–DNA contacts. The *in vivo* and *in vitro* studies together with sequence analysis of the eight functionally active AraR *B. subtilis* boxes indicated that bases G_9, A_{11} and T_{16} are crucial for AraR binding, and that A_1 and T_6 play also an important role. Furthermore, Position 5 required a purine (Pu) for functionality *in vivo*, and sequence analysis suggested that the corresponding mirror base (Position 12) in the other arm of the palindrome is always a pyrimidine (Py). An alignment of all putative AraR-binding sequences based on a search of the consensus 5′-ATTTGTACGTACAAAT-3′ in genomes of bacteria from the *Bacillus*/*Clostridium* group that contain AraR orthologues also highlighted the majority of the invariable positions: Pu_5, T_6 and the correspondent mirror A_{11}, Py_{12} and G_9 at the centre of the palindrome (Supplementary Figure 1).

Mutation M34T partially restores AraR binding to mutated DNA operators both *in vivo* and *in vitro*

In order to isolate AraR mutants that could suppress the loss of regulation caused by the single nucleotide substitutions in the *ara* boxes, an *in vivo* screening method was developed (see Materials and Methods section). Briefly, random mutagenesis of the 5′-end of the *araR* allele was performed by PCR and the resulting library of plasmids carrying the mutated alleles was used to transform *B. subtilis* strains, allowing its integration at the non-essential *thrC* locus via a double-crossover event. The receptor strains possessed a Δ*araR* Δ*amyE::OR$_A$-lacZ* background, and carried different mutations in the OR_A operator sequence (Table 1). The constitutive expression of *lacZ* due to the inability of the wild-type AraR to bind the mutated operator leads to a Lac^+ phenotype in the absence of inducer. However, if the integrated mutant *araR* allele encodes a protein that suppressed the deleterious effect of the operator mutation

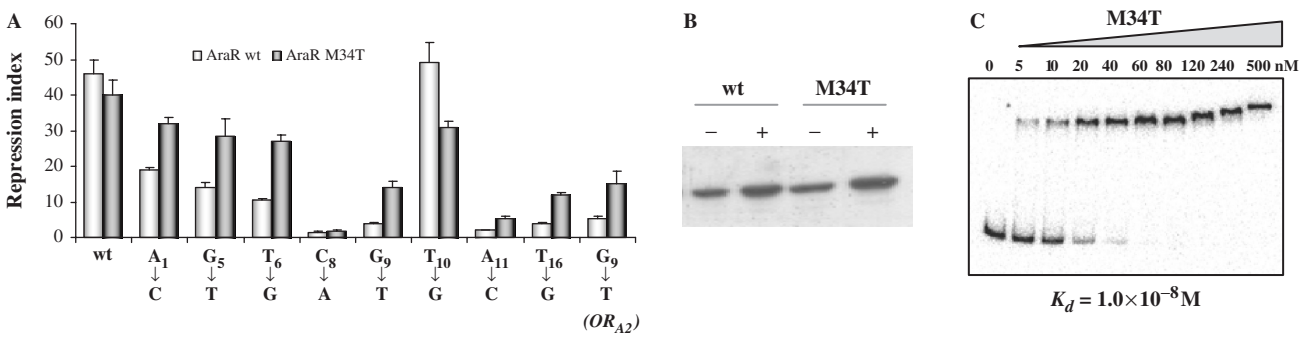


Figure 6. Characterization of suppressor mutant M34T. (**A**) Comparison of the repression exerted *in vivo* by AraR wild type or mutant M34T over *ara* boxes. β-Galactosidase activities were determined in *B. subtilis* strains carrying the different *araA–lacZ* fusions integrated at the *amyE* locus and the wild-type or M34T *araR* allele integrated at the *thrC* locus, grown in inducing and non-inducing conditions. The bars represent the repression index obtained for wild-type AraR (white bars) or variant M34T (grey bars) (Figure 2). Values are the average of three independent experiments, each assayed in duplicate. Error bars represent the SD. (**B**) Intracellular accumulation of AraR wild type and mutant M34T by western immunoblot analysis from cultures grown in the absence (−) or presence (+) of inducer (Figure 2). (**C**) Binding of AraR M34T to *araABDLMNPQ-abfA* promoter (operators OR_{A1}–OR_{A2}) in EMSA. The indicated concentrations of AraR mutant M34T were used in the binding reactions. The derived K_d is shown below (Figure 3).

a Lac$^-$ phenotype is displayed indicating recovery of regulation. Thus, we screened for transformants with decreased β-galactosidase production in the absence of arabinose.

One transformant of strain IQB533, containing the fusion $OR_{A1(T6\rightarrow G)}$–*lacZ*, displayed a gain-of-function phenotype, and the sequencing of the *araR* allele revealed the substitution M34T, located in the HTH motif of the protein. To determine if this effect was specific or also affected AraR binding to the other mutated promoters, the allele was integrated at the *thrC* locus of the corresponding *B. subtilis* strains and β-galactosidase activities were measured (Figure 6A). Interestingly, the repression exerted by mutant M34T was higher than that exerted by wild-type AraR in almost all operators, in particular with mutant boxes $G_9\rightarrow T$ (both in OR_{A1} and OR_{A2}), $T_6\rightarrow G$, $A_{11}\rightarrow C$ and $T_{16}\rightarrow G$. Since a higher level of intracellular accumulation of this AraR variant could explain this phenotype, we determined the accumulation of the wild-type AraR and mutant M34T in strains IQB572 and IQB583, respectively (Table 1). The observed cellular levels of protein were similar in both strains (Figure 6B), indicating that the phenotype displayed by M34T was not due to increased concentration of protein. In fact, EMSA assays performed as described above showed that the mutant displays an increased affinity to the *ara* operon promoter probe, with an apparent K_d of 1.0×10^{-8} M (Figure 6C), which is almost 4-fold lower than that of the wild-type protein ($K_d = 3.9\times10^{-8}$ M). The substitution M34T is located in the first helix of the winged HTH motif (Figure 1A). According to a study on protein–DNA interactions based on structures of 129 complexes (32), threonine is responsible for a far larger number of protein–DNA bonds than methionine, although almost all are made with the DNA backbone and not the bases. This is in agreement with the results we obtained, that show an increase in the repression exerted over all mutated boxes, suggesting an increased DNA affinity of M34T through non-specific contacts.

CONCLUDING REMARKS

Previous studies have mapped the functional domains of AraR and characterized the C-terminal region involved in effector binding and dimerization (8). In this work, we focused on two additional and crucial components of the transcription process, the DBD and *cis*-acting elements. Guided by molecular modelling combined with multiple primary sequence alignment of AraR orthologues and GntR family members, we identified amino acids potentially involved in DNA binding. The effect of their substitution was analysed *in vivo* and *in vitro* and revealed key residues necessary for the regulatory activity. In addition, important bases for AraR–DNA interactions in both arms of the palindromic operator sequences were also identified. We obtained both expected and unexpected results highlighting the uniqueness of protein–DNA interactions in each particular system. A future determination of the structure of AraR, in its unbound form or in complex with the inducer or DNA, would allow a more detailed analysis of the mechanism by which AraR binds its cognate operator sequences and how the conformational change triggered by the binding of arabinose prevents this interaction.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was partially supported by grants POCTI/BME/36164/00 and POCI/BIA-MIC/61140/04 from Fundação para a Ciência e Tecnologia (FCT) and FEDER to I.S.-N. I.S.F. is the holder of Ph.D. fellowship SFRH/BD/5233/01 from FCT. Funding to pay the Open Access publication charges for this article was provided by Fundação para a Ciência e Tecnologia.

Conflict of interest statement. None declared.

REFERENCES

1. Sá-Nogueira,I., Nogueira,T.V., Soares,S. and de Lencastre,H. (1997) The *Bacillus subtilis* L-arabinose (ara) operon: nucleotide sequence, genetic organization and expression. *Microbiology*, **143** (Pt 3), 957–969.
2. Sá-Nogueira,I. and Ramos,S.S. (1997) Cloning, functional analysis, and transcriptional regulation of the *Bacillus subtilis araE* gene involved in L-arabinose utilization. *J. Bacteriol.*, **179**, 7705–7711.
3. Sá-Nogueira,I. and Mota,L.J. (1997) Negative regulation of L-arabinose metabolism in *Bacillus subtilis*: characterization of the *araR* (*araC*) gene. *J. Bacteriol.*, **179**, 1598–1608.
4. Raposo,M.P., Inácio,J.M., Mota,L.J. and de Sá-Nogueira,I. (2004) Transcriptional regulation of genes encoding arabinan-degrading enzymes in *Bacillus subtilis*. *J. Bacteriol.*, **186**, 1287–1296.
5. Krispin,O. and Allmansberger,R. (1998) The *Bacillus subtilis* AraE protein displays a broad substrate specificity for several different sugars. *J. Bacteriol.*, **180**, 3250–3252.
6. Mota,L.J., Tavares,P. and Sá-Nogueira,I. (1999) Mode of action of AraR, the key regulator of L-arabinose metabolism in *Bacillus subtilis*. *Mol. Microbiol.*, **33**, 476–489.
7. Mota,L.J., Sarmento,L.M. and de Sá-Nogueira,I. (2001) Control of the arabinose regulon in *Bacillus subtilis* by AraR in vivo: crucial roles of operators, cooperativity, and DNA looping. *J. Bacteriol.*, **183**, 4190–4201.
8. Franco,I.S., Mota,L.J., Soares,C.M. and de Sá-Nogueira,I. (2006) Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. *J. Bacteriol.*, **188**, 3024–3036.
9. Haydon,D.J. and Guest,J.R. (1991) A new family of bacterial regulatory proteins. *FEMS Microbiol. Lett.*, **63**, 291–295.
10. Weickert,M.J. and Adhya,S. (1992) A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.*, **267**, 15869–15874.
11. Rigali,S., Derouaux,A., Giannotta,F. and Dusart,J. (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR and YtrA subfamilies. *J. Biol. Chem.*, **277**, 12507–12515.
12. Lee,M.H., Scherer,M., Rigali,S. and Golden,J.W. (2003) PlmA, a new member of the GntR family, has plasmid maintenance functions in *Anabaena* sp. strain PCC 7120. *J. Bacteriol.*, **185**, 4315–4325.
13. Geer,L.Y., Domrachev,M., Lipman,D.J. and Bryant,S.H. (2002) CDART: protein homology by domain architecture. *Genome Res.*, **12**, 1619–1623.
14. Clark,K.L., Halay,E.D., Lai,E. and Burley,S.K. (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, **364**, 412–420.
15. Gajiwala,K.S. and Burley,S.K. (2000) Winged helix proteins. *Curr. Opin. Struct. Biol.*, **10**, 110–116.
16. Luscombe,N.M. and Thornton,J.M. (2002) Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. *J. Mol. Biol.*, **320**, 991–1009.
17. Miller,J. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NYC, USA.
18. Pascal,M., Kunst,F., Lepesant,J.A. and Dedonder,R. (1971) Characterization of two sucrase activities in *Bacillus subtilis* Marburg. *Biochimie*, **53**, 1059–1066.
19. Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.
20. Sambrook,J., Fritsch,E. and Maniatis,T. (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York, USA.
21. Guerout-Fleury,A.M., Frandsen,N. and Stragier,P. (1996) Plasmids for ectopic integration in *Bacillus subtilis*. *Gene*, **180**, 57–61.
22. Leung,D., Chen,E. and Goeddel,D.V. (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique*, **1**, 11–15.
23. Bera,A.K., Zhu,J., Zalkin,H. and Smith,J.L. (2003) Functional dissection of the *Bacillus subtilis pur* operator site. *J. Bacteriol.*, **185**, 4099–4109.
24. van Aalten,D.M., DiRusso,C.C. and Knudsen,J. (2001) The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. *EMBO J.*, **20**, 2041–2050.
25. Riggs,A.D., Suzuki,H. and Bourgeois,S. (1970) Lac repressor-operator interaction. I. Equilibrium studies. *J. Mol. Biol.*, **48**, 67–83.
26. Raman,N., Black,P.N. and DiRusso,C.C. (1997) Characterization of the fatty acid-responsive transcription factor FadR: biochemical and genetic analyses of the native conformation and functional domains. *J. Biol. Chem.*, **272**, 30645–30650.
27. van Aalten,D.M., DiRusso,C.C., Knudsen,J. and Wierenga,R.K. (2000) Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold. *EMBO J.*, **19**, 5167–5177.
28. Xu,Y., Heath,R.J., Li,Z., Rock,C.O. and White,S.W. (2001) The FadR.DNA complex. Transcriptional control of fatty acid metabolism in *Escherichia coli*. *J. Biol. Chem.*, **276**, 17373–17379.
29. Yoshida,K., Fujita,Y. and Sarai,A. (1993) Missense mutations in the *Bacillus subtilis gnt* repressor that diminish operator binding ability. *J. Mol. Biol.*, **231**, 167–174.
30. Rodionov,D.A., Mironov,A.A. and Gelfand,M.S. (2001) Transcriptional regulation of pentose utilisation systems in the *Bacillus*/*Clostridium* group of bacteria. *FEMS Microbiol. Lett.*, **205**, 305–314.
31. Yoshida,K., Yamamoto,Y., Omae,K., Yamamoto,M. and Fujita,Y. (2002) Identification of two myo-inositol transporter genes of *Bacillus subtilis*. *J. Bacteriol.*, **184**, 983–991.
32. Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
33. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.