Unique DNA Binding Specificity of the Binuclear Zinc AlcR Activator of the Ethanol Utilization Pathway in Aspergillus nidulans*

(Received for publication, October 28, 1998, and in revised form, January 25, 1999)

Igor Nikolaev†, François Lenouvel‡, and Béatrice Felenbok¶
From the Institut de Génétique et Microbiologie, Unité Mixte de Recherche CNRS no 8621, Université Paris-Sud XI, Bâtiment 409, Centre Universitaire d’Orsay, F-91405 Orsay Cedex, France

AlcR is the transcriptional activator in Aspergillus nidulans, necessary for the induction of the alc gene cluster. It belongs to the Zn$_2$Cys$_6$ zinc cluster protein family, but contains some striking differences compared with other proteins of this group. In this report, we show that no dimerization element is present in the entire AlcR protein which occurs in solution as a monomer and binds also to its cognate sites as a monomer. Another important feature of AlcR is its unique specificity for single sites occurring naturally as inverted or direct repeats and sharing a common motif, 5’-(T/A)GCGG-3’. Like most other Zn$_2$Cys$_6$ proteins, AlcR contacts directly with the CGG triplet and, in addition, the upstream adjacent guanine is required for high affinity binding. We also establish that the flanking regions outside the core play an essential role in tight binding. From our in vitro analysis, we propose an optimal AlcR-binding site which is 5’-PuNGCGG-AT rich 3’.

The Aspergillus nidulans protein, AlcR, activates transcription of genes required for the oxidation of ethanol and other alc clustered genes whose functions are currently being studied (1, 2). Transcriptional regulation is mediated by the binding of AlcR to different cognate sites, encompassing the same consensus core, 5’-(T/A)GCGG-3’, organized as direct and inverted repeats, located upstream of these genes (3–6). AlcR is a fungal transcription factor of the zinc binuclear cluster family, containing six cysteines and two zinc atoms (C6 protein class) (7–9). However, AlcR is distinguishable from other proteins in this family by its structural and binding properties. It has an asymmetric structure resulting from an additional 16 residues between the third and the fourth cysteines compared with 6–8 residues usually found in the other proteins of this class (7). Furthermore, in this loop, the proline residue conserved in all the zinc binuclear cluster domains and which was shown to be important for zinc binding, as stated for GAL4 (10) and more recently for PrnA (11), is not present in AlcR. However, these differences have no effect on the distance between the two zinc atoms and between zinc and sulfhydril ligands of cysteine (12) which are similar to those found in the solved structure of GAL4, PPR1, and PUT3 (13–16). Another important feature of AlcR is the absence of a coiled-coil dimerization region following the C6 zinc cluster in the other proteins (5). Two stretches of leucine heptad repeats, downstream of the zinc cluster, are interrupted by several proline residues known to disrupt a continuous a-helix.

NMR studies have shown that the isolated AlcR zinc binuclear cluster domain (residues 1–60) binds to a single copy site as a monomer, albeit with a low affinity (9). More recently, binding experiments performed with a longer AlcR protein (1–197), containing the region downstream of the zinc cluster, have clearly indicated that one AlcR molecule binds with high affinity to DNA single sites (5 x 10$^{-9}$ M) (5, 6). Furthermore, using a transcription-translation assay, we have shown that no dimerization sequence is present up to residue 197 which comprises one-fourth of the protein length (5).

In this work, we wanted to address several questions. First, it was important to determine if any functional dimerization element other than those generally found in the proteins of the C6 class was present in AlcR, downstream of residue 197. Second, since the AlcR consensus core includes both the CGG triplet found to interact with the other well defined zinc binuclear cluster proteins such as GAL4, PPR1, PUT3 (13–16), and the GCG triplet shown to interact with a few other C6 proteins (17), the mode of recognition of AlcR toward its DNA sites was investigated. Third, it appears that AlcR specificity toward its DNA target is different from the other proteins of the C6 class, thus the role of the flanking regions outside the consensus core was determined. Since AlcR natural targets are organized as direct or inverted repeats, the size and the composition of the spacer are also of special interest in the in vitro study of AlcR binding.

In order to answer these questions, a functional analysis of the putative dimerization regions of AlcR has been performed as well as a systematic study by site-directed mutagenesis of AlcR binding to direct or inverted repeat physiological targets and to an artificial single site.

EXPERIMENTAL PROCEDURES

Strains, Plasmids and Phages—Escherichia coli strains CSH50: Δ(proLac) F’ (proABlacQ) AM15, traDS6 was used in phage immunity tests. Wild type phage λ was employed for immunity tests. The pCIR1–7 plasmids were obtained from the pC132 vector (18) which harbors a CI rop fusion by replacing the rop gene with the alcR regions of interest. The different AlcR domains were amplified by polymerase chain reaction using specific oligonucleotides containing SaI and BamHI sites at their 5’ and 3’ termini, respectively. The polymerase chain reaction products were cut by SaI and BamHI and cloned in the pC132 plasmid treated with the same enzymes to excise the rop sequence. To generate plasmids pCIR8 and pCIR9 containing the alcR fusion with the ci repressor and the rop genes, the alcR fragment was

* This work was supported in part by Centre National de la Recherche Scientifique Grant URA D 2225, by the Université Paris-Sud XI, and by European Communities Contract B044-CT96-0535. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
† Supported by grants from NATO and as an Invited Professor.
‡ Recipient of doctoral fellowship from the Ministère de l’Éducation Nationale, de l’Enseignement Supérieur et de la Recherche of the French Government. Present address: Molecular Genetics of Industrial Microorganisms, Wageningen Agricultural University, Dreijenlaan 2, 6703 HA Wageningen, The Netherlands.
¶ To whom correspondence should be addressed. Tel.: 33-1-69-15-63-28; Fax: 33-1-69-15-78-08; E-mail: felenbok@igmors.u-psud.fr.

This paper is available online at http://www.jbc.org
amplified by polymerase chain reaction primed from specific oligonucleotides containing SacI sites at both ends and cloned into the pC132 plasmid cut by SalI. In all these constructs the AlcR domains were fused in-frame with the λ phase CI repressor NH₂-terminal domain at the 5’ end and the β-galactosidase α-peptide gene at the 3’ end. The AlcR coding sequences were separated from the α-peptide by an amber codon introduced during amplification to allow for synthesis of a bipartite (CI:AlcR), tripartite (CI:AlcR:α-peptide), or tetrapartite (CIAlcR:Rop:α-peptide) fusion protein depending on the suppressor background of the host. The supE (71.18) strain bearing fusion plasmids forms blue colonies on 5-bromo-4-chloro-3-indoyl-β-D-galactoside indicator plates. Bacterial cells transformed by plasmids were assayed by spot tests for sensitivity to λ phages at different concentrations on lawns of transformed bacteria.

The plasmid expressing AlcR(1–197) tagged with 6xHis was constructed as described previously (5). The AlcR protein expressed was partially purified on a Ni²⁺/nitrilotriacetic acid-agarose column at up to 20% of homogeneity and used for electrophoretic mobility shift assays.

Electrophoretic Mobility Shift Assays—Oligonucleotide probes containing either direct repeat targets a or c or inverted repeat target b from the alcA promoter or an artificial single copy site sc and its mutated derivatives were used in most gel shift experiments. Sequence of these probes, as present on the top strand of the alcA promoter sequence, are given below: a, 5'-CCCTCTTCCGGGGCCGAC-3'; b, 5'-GTGCGG-3'; c, 5'-CTTCTTCCGGGGCCGAC-3'; e, 5'-CTTCTTCCGGGGCCGAC-3'. The conserved repeats are marked in boldface type. The sequence of a single copy probe site sc and its mutated variants is indicated in Table II. Probes with a modified spacer region (underlined below) have the following sequence: inverted repeats: N2, 5'-ATGGCATGGTAGCTTTCGGCAGAACC-3'; N2, 5'-ATGGCATGGTAGCTTTCGGCAGAACC-3'.

The plasmid expressing AlcR(1–197) tagged with 6xHis was calibrated to yield approximately 30–50% of bound protein, was introduced at the carboxyl terminus of AlcR sequenced before gel filtration chromatography. The AlcR(1–197) protein migrated with an apparent molecular mass of 18 kDa (data not shown), a value close to that expected for a monomer (22 kDa). Nevertheless, the truncated protein could be lacking a downstream dimerization domain. In order to determine if AlcR regions other than those usually found in the C6 class proteins could drive dimerization of AlcR, another method based on the property of the bacteriophage λ CI repressor to function as a dimer was used (20). When a dimerization element is fused to the NH₂-terminal DNA-binding domain of the λ CI repressor, E. coli cells are immune to superinfection by λ phage, while cells that contain the CI DNA-binding domain alone are sensitive. This system was used previously to analyze the residues involved in the leucine zipper from the yeast activator GCN4 (20), in TAT homodimerization (21), Myc/Max heterodimer formation (22), and dimerization of the NisA zinc cluster protein (23). Gene fusions between the NH₂-terminal domain of λ CI repressor and overlapping AlcR sequences covering the whole protein (1–821), including the zinc cluster domain (1–59), were performed (as described under “Experimental Procedures”). None of the AlcR sequences tested conferred immunity to λ infection when transformed into E. coli, implying that no dimerization element was present (Table 1).

One could argue that the lack of dimerization could be the result of a possible insolubility or misfolding of the chimeric λ CI-AlcR protein. This hypothesis was ruled out by assaying the β-galactosidase activity driven by all the fusion constructs (see “Experimental Procedures”). In order to ensure that AlcR sequences did not prevent per se, dimerization of a CI repressor, a control dimerization element, originating from the Rop protein, was introduced at the carboxyl terminus of AlcR sequences (pCIR8, pCIR9). Results presented in Table I show clearly that this Rop sequence is able to drive a CI dimerization when fused at the carboxyl-terminal region of AlcR, indicating that AlcR sequences did not prevent dimerization. Another hypothesis could be that the AlcR dimerization element consists of separate domains, comprising interacting α helical regions. The only possibility, according to computer prediction analysis of AlcR (24) and preliminary results of AlcR NMR studies,¹ would be the interaction of two regions, one of 16 residues, located near the C-terminal part of AlcR, as shown in the figures. In all cases, only the top strand of the oligonucleotide is shown. The binding reactions were analyzed by gel electrophoretic mobility shift assays as described previously (5) and were quantified with a PhosphorImager (Molecular Dynamics). In experiments determining the relative binding efficiencies of oligonucleotides containing sc or mutated sequences, the protein concentration of AlcR(1–197) was calibrated to yield approximately 30–50% of bound probe. Binding affinities were relative to the binding of AlcR(1–197) to an artificial probe sc normalized to 1 or 100.

Footprint Analysis—Methylation interference and denaturation interference footprinting assays were carried out as described earlier (4). Generally, 10⁵ cpm of end-labeled single strand DNA probe were annealed with the complementary, non-labeled strand and then subjected to chemical modifications either with dimethyl sulfate, for partial methylation of guanines, or piperidine formate, for partial depurination, according to the procedure of Maxam and Gilbert (19). Briefly, the modified DNA probes were incubated with sufficient partially purified AlcR(1–197) protein (300–500 ng) to bind approximately 30–50% of the probe. The protein bound and free DNAs were separated by preparative gel mobility shift technique and electroeluted from gel slices. Recovered DNAs were treated with 1 x piperidine and resolved on a 16% polyacrylamide gel containing 8 M urea.

Footprint Analysis—Methylation interference and denaturation interference footprinting assays were carried out as described earlier (4). Generally, 10⁵ cpm of end-labeled single strand DNA probe were annealed with the complementary, non-labeled strand and then subjected to chemical modifications either with dimethyl sulfate, for partial methylation of guanines, or piperidine formate, for partial depurination, according to the procedure of Maxam and Gilbert (19). Briefly, the modified DNA probes were incubated with sufficient partially purified AlcR(1–197) protein (300–500 ng) to bind approximately 30–50% of the probe. The protein bound and free DNAs were separated by preparative gel mobility shift technique and electroeluted from gel slices. Recovered DNAs were treated with 1 x piperidine and resolved on a 16% polyacrylamide gel containing 8 M urea.

Results listed in Table I show clearly that this Rop sequence is able to drive a CI dimerization when fused at the carboxyl-terminal region of AlcR, indicating that AlcR sequences did not prevent dimerization. Another hypothesis could be that the AlcR dimerization element consists of separate domains, comprising interacting α helical regions. The only possibility, according to computer prediction analysis of AlcR (24) and preliminary results of AlcR NMR studies,¹ would be the interaction of two regions, one of 16 residues between the third and fourth cysteines and the second region downstream of the zinc cluster at residues 310 to 343. Results presented in Table I (pCIR4) argue against this assumption. In fact, the AlcR region from residues 1 to 367 does not confer immunity to λ infection whereas the AlcR-Rop fusion does. Therefore, the absence of immunity, whatever the alcR

| Plasmid | Protein domain fused to DBD of CI λ repressor | Test for λ phage infection* |
|---------|-----------------------------------------------|-----------------------------|
| None    |                                               | Sensitive                   |
| pC1     |                                               | Sensitive                   |
| pC132   | Rop                                           | Immune                      |
| pCIC1R1 | AlcR(1–59)                                    | Immune                      |
| pCIC1R2 | AlcR(102–191)                                  | Sensitive                   |
| pCIC1R3 | AlcR(50–367)                                   | Sensitive                   |
| pCIC1R4 | AlcR(1–367)                                    | Sensitive                   |
| pCIC1R5 | AlcR(372–584)                                  | Sensitive                   |
| pCIC1R6 | AlcR(585–821)                                  | Sensitive                   |
| pCIC1R7 | AlcR(1–821)                                    | Sensitive                   |
| pCIC1R8 | AlcR(102–191) Rop                             | Immune                      |
| pCIC1R9 | AlcR(1–367) Rop                               | Immune                      |

* A phages of wild type were used in spot tests at concentrations of 10⁻⁵ to 10⁻⁶ phage per spot, on lawns of CSH50 E. coli cells harboring different plasmid constructs (18).

1 R. Cerdan, F. Penin, F. Lenouvel, B. Felenbok, and E. Gutett, unpublished results.
sequence fused to λ CI repressor, indicates that under these conditions the AlcR dimerization sequence, if any, is either not functional or not strong enough to promote in vivo formation of dimers of CI repressor.

To obtain additional evidence that AlcR does not form dimers, at least in solution, we compared the dimerization of AlcR and NirA by cross-linking each protein in solution. NirA has been recently shown to form stable dimers (23). A full-length AlcR protein and a truncated NirA(1–222) version containing its dimerization interface were translated in vitro and after treatment with glutaraldehyde were analyzed by SDS-gel electrophoresis. As shown in Fig. 1, the majority of NirA(1–222) dimerizes whereas AlcR(1–821) stays monomeric in solution even in conditions when the translated product is increased 7-fold in the reaction.

Mode of AlcR Interaction with Its Cognate Targets—Our previous in vitro DNA binding studies using alcA physiological repeated sites have shown the formation of two complexes (a fast migrating complex CI and a slow migrating complex CII) with symmetric sites, which correspond to the binding of one fast migrating complex CI and a slow migrating complex CII) repeated sites have shown the formation of two complexes (an previous studies (5).

For both asymmetric targets a and c, only a slight reduction of signal at positions encompassing the consensus motif 5’-GCGG-3’ and the adjacent A (position 6) was detected (Fig. 2, A and B). The equal reduction of band intensities for both repeats is indicative of a random occupancy of binding sites which results in formation of one retarded complex, CI. Our present results might seem contradictory to the previous footprinting data which demonstrated that only one of the two binding sites, namely c1, can be occupied by AlcR. However, the oligonucleotide probe utilized earlier was 6 bp shorter at the 5’ end than the one we used in the current studies. Apparently, the limited size of flanking regions impaired the AlcR binding to the adjacent site c1.

Similarly, AlcR exhibits the same manner of site recognition, when fixed on the symmetric target, giving rise to the fast migrating complex, CI. Only a slight interference is noticeable by both methylation and missing-contact footprint analyses (Fig. 2D). Thus, no site preference for binding of the first AlcR molecule to either direct or inverted repeat sites in the alcA promoter was observed. A depurination footprint pattern obtained with complex CII revealed a strong interference of all the G residues within the consensus motif in both strands. In addition, two central A residues separating inverted repeats exhibited strong interference (Fig. 2D). Methylation of two G residues at positions 2 and 4 in both strands interfered strongly with AlcR binding, while methylation of the adjacent G (position 5) had weaker effects. These results indicate that AlcR interacts mainly in the major groove, although minor groove contacts exist. For example, contacts with N3 of A in the spacer sequence in the symmetric repeat b, as shown by methylation footprinting data, occur in the minor groove (Fig. 2D). Within each recognition motif AlcR makes direct, sequence-specific contacts with bases on the same strand, since methylation of the N-7 position of G (position 3) in the C-rich strand of the consensus core resulted in weak, if any, interference.

Footprint experiments performed with an artificial single copy site (sc), designed after disruption of one of the repeats in the symmetric target, suggest that the AlcR molecule establishes direct base-specific contacts within the conserved repeat and does not interact with neighboring nucleotides (Fig. 2C). The pattern of depurination interference appeared to be similar for all types of targets (inverted or direct repeat or a single copy site), indicating that AlcR uses a unique strategy for DNA recognition.

AlcR Interacts Directly with the CGG Triplet—Most of the zinc binuclear cluster proteins have been shown to interact directly with CGG triplets present at the ends of their cognate targets, whereas some activators of this class are able to recognize the CGC triplet (25). According to footprinting data, the AlcR-binding site represents a combination of both triplets found on opposite strands. To distinguish among these various putative sites as well as to examine the contribution of each interacting base, we tested the effect of base substitutions within the consensus motif using saturation mutagenesis. AlcR-binding affinities for all mutant-binding sites were estimated from gel band shift experiments and compared with that of the single copy probe sc. Changes of any base within the CGG triplet severely reduced the affinity for AlcR. Of these bases, the G residues at positions 4 and 5 were the most sensitive to change, since substitution with any of the other three bases resulted in a complete loss of binding (Table II). Mutation of C in position 3 resulted also in a drastic decrease of AlcR binding (50–100). A large decrease (10–20 fold) in binding activity was observed by mutations in position 2. Although this reduction is smaller than those found when mutations are introduced in the terminal CGG triplet, it is substantial compared with the effect of substitutions at position 1. The highly conserved T at this position makes a negligible contribution to AlcR binding affinity, since any nucleotide gives reasonably strong binding. Consistent with the footprinting data, these results confirm the absence of essential base-specific contacts at this position. On the whole, our results suggest that, like other proteins of the zinc binuclear cluster family, AlcR recognizes the CGG triplet.
by establishing crucial contacts, presumably, in the major groove. However, unlike the other proteins of this C6 class, position 2 occupied by G is also essential for AlcR high affinity binding. The first nucleotide seems to be important for tight binding rather than for a specific interaction with the DNA.

The Distance between Two Sites Specifies Binding of Two AlcR Molecules—All the naturally occurring AlcR inverted repeat targets identified in the \textit{alcA} promoters display conserved central bases, APu (2 bp)\textsuperscript{2} separating two inverted motifs, whereas within direct repeats the spacing sequence varies greatly (3–8 bp) (Fig. 3). Moreover, in symmetric sites, these bases were shown to be involved in direct contacts with AlcR (see Fig. 2). We wanted to know if the composition and spacing length are critical for AlcR binding. Since removal of these bases by the missing-contact technique interferes with AlcR binding, it seemed important to estimate the contribution of the inner 2-bp spacer in symmetric sites, for protein binding.

Substitution of the spacer sequence in the \textit{alcA} symmetric site from AA to GC bases which is never encountered in natural palindromic \textit{alc} sites, reduced AlcR binding by 3–5-fold (Fig. 4A). This is in agreement with results of the footprinting analysis (Fig. 2D) showing that contacts established within the spacing region contribute to some extent to specific binding and, hence, interaction extends beyond the consensus core.

\textsuperscript{2} The abbreviation used is: bp, base pair(s).
**DNA Binding Specificity of AlcR Binuclear Cluster**

**TABLE II**

| Mutation | Sequence of the probe | % of binding affinity relative to the probe sc |
|----------|-----------------------|---------------------------------------------|
| sc       | 5'-GATGCATCGCGGAAACGACGAGGG-3' | 100 |
| 1A       | 5'-GATGCAGCGGAAACGACGAGGG-3' | 100 |
| 1G       | 5'-GATGCAGCGGAAACGACGAGGG-3' | 50 |
| 1C       | 5'-GATGCACGCGGAAACGACGAGGG-3' | 70 |
| 2A       | 5'-GATGCATAACGACGAGGG-3' | 10 |
| 2T       | 5'-GATGCATCTCGGAAACGACGAGGG-3' | 8 |
| 2C       | 5'-GATGCATCGGAAACGACGAGGG-3' | 5 |
| 3A       | 5'-GATGCATAAGCAGCAGCAGGAGGG-3' | 2 |
| 3G       | 5'-GATGCATCGGAAACGACGAGGG-3' | 2 |
| 3T       | 5'-GATGCATACTGGAAACGACGAGGG-3' | 3 |
| 4A       | 5'-GATGCATGCAAGAAGCAGCAGGAGGG-3' | <1 |
| 4T       | 5'-GATGCATACGAAGCAGCAGGAGGG-3' | <1 |
| 4C       | 5'-GATGCATCGGAAACGACGAGGG-3' | <1 |
| 5A       | 5'-GATGCATCGGAAACGACGAGGG-3' | <1 |
| 5T       | 5'-GATGCATCGGAAACGACGAGGG-3' | <1 |
| 5C       | 5'-GATGCATCGGAAACGACGAGGG-3' | <1 |

* a Mutated probes are named according to the corresponding substitutions in the consensus sequence 5'-TGCGG-3' (underlined). Numbers correspond to base positions where mutations were made and letters correspond to nucleotides introduced at these positions. Sequence of the nonmutated probe sc corresponds to an artificial single copy site designed after mutation of the inverted repeat target b. Only relevant sequences in one strand containing the consensus core in the 5' to 3' direction are presented. The mutated bases are indicated by boldface type.

b Binding affinities for each probe were estimated from mobility shift assays as described under "Experimental Procedures." The percentage of the free probe. The ratio obtained for the probe sc was set to 100% and ratios for mutant probes are relative to that of the corresponding free probe.

To test whether the natural spacing between inverted repeats is a prerequisite for binding, we assayed AlcR binding to mutant probes with increasing spacer lengths (Fig. 4). As seen in the model summarized in Fig. 5, the 2-bp spacing between inverted repeats in the alcA promoter resulted in positioning of the cores on the opposite sides of the DNA helix. In this case, two AlcR molecules are presumed to be in a head to head orientation. Surprisingly, increasing the spacer length just by one nucleotide (spacer N3) completely abolished the formation of the slow migrating complex CII (containing two AlcR molecules), suggesting that one AlcR molecule is bound with an affinity similar to that of a single site (Fig. 4). Footprinting analysis performed with the CI complex (containing one AlcR molecule) showed a random occupancy of the two inverted repeats (results not shown). A similar pattern of binding was also observed when the spacer region was gradually increased to 7 bp, thereby placing the two interacting CCG triplets on the same face of the double helix. The absence of simultaneous binding of two AlcR molecules to such modified palindromes might be due to the fact that fixation of the first AlcR molecule on the DNA induces a conformational change in the DNA, preventing the binding of the second AlcR molecule. Another hypothesis, implying the interaction between two AlcR molecules, seems less convincing. If such an interaction did exist, a strong cooperativity of AlcR binding would have been observed, which was not the case (5, 24). More likely, the phasing of the sites is important for binding of two AlcR molecules. Increasing the spacer length to 12 bp, corresponding to one additional turn of the DNA helix, did not restore normal binding of AlcR. We found that the protein did indeed bind to the second site but with an affinity much lower than with the wild type probe (Fig. 4A, spacer N12). This decrease in affinity could be due to changes in spacing rather than in adjacent sequences to the consensus core, which are unchanged. The same decrease was already observed when changing the two nucleotides of the spacer (Fig. 4A, spacer GC). These results are in agreement with data presented below, showing a general role of flanking regions in high binding affinity. In contrast, a further increase in the distance between triplets by one-half turn of helix (spacer N-17), results in the formation of the slow migrating complex CII with a similar range of affinity characteristic of the wild type probe (3-fold lower). In that case, the two repeated sites are placed on the same face of the double helix but one turn apart.

Natural spacing between physiologically functional direct repeats in the alcA promoter varies from 7 to 8 bp (Fig. 3), which corresponds to a distance between the CCG triplets of 9 and 10 bp, respectively. Hence, both sites partly lie on the same face of the DNA helix (Fig. 5). Assuming a head to tail orientation of AlcR for this case, such a positioning of direct repeated sites on DNA obviously does not favor simultaneous occupancy of both sites by AlcR, as observed previously (5, 6) and in Fig. 4B (spacer N-10). Introducing an additional 3 bp into the spacer region of target e removes the second CCG triplet on the opposite side of the double helix (spacer N-13). Indeed, the second site is then accessible for AlcR high affinity binding.
resulting in the formation of the complex CII. Therefore, with both inverted or direct repeats, the intervening sequence is involved in the binding of two AlcR molecules.

**Role of the Flanking Regions Outside the Consensus Core**—Sequence alignment of AlcR symmetric sites found in the AlcR regulated promoters, allowed the identification of conserved base pairs outside the consensus motif (Fig. 6A). Thus, the 5'-proximal position is characterized by the presence of a purine preceded by C or A. It appears that the 5'-proximal purine is more important for high affinity binding, since its replacement results in a 10-fold decrease in binding (Fig. 6B). This indicates that, in addition to the conserved A immediately downstream of the consensus motif, shown to interact directly with AlcR, the 5'-proximal bp contributes significantly to binding. This result is consistent with the observation that AlcR high affinity binding sites in direct repeat targets of the alcA promoter are often flanked by the same conserved nucleotides (Fig. 6C). In contrast, sites with low binding affinity are followed by a G-rich stretch (Fig. 6C). In order to simplify the analysis, nonphysiological artificial single copy sites were further investigated. Mutagenesis in the 5'- and 3'-flanking regions was carried out. In Fig. 6D only the mutated oligonucleotides relevant to AlcR binding are shown. In fact, the purine 5'-proximal to the binding site plays an active role in DNA recognition by AlcR. Substitution of A by T reduced the affinity 5-fold, whereas mutating the preceding position −2 (C to G) had little effect. Apart from the highly conserved A 3' to the consensus core which, as expected from footprinting analysis, contributes to binding, mutagenesis of any other base in the 3'-flanking region did not lead to any significant effect (results not shown). However, while no single mutation in this region is essential, binding of AlcR is dependent on the overall composition of the sequence 3'-proximal to the site. Changing the box composition of nucleotides from AT-rich to G-rich results in a noticeable decrease (3–9-fold) of AlcR binding (Fig. 6D). Based on these data, it can now be explained why some of the AlcR-binding sites, as for example, the direct repeated sites d in the alcR promoter (Fig. 3) are not active in vivo.3

**DISCUSSION**

The activator AlcR belongs to the zinc binuclear cluster family that includes a number of other fungal transcription factors. These factors share common regions essential for DNA binding (13, 17, 25, 26). They bind as homodimers to their DNA cognate targets, organized either as inverted or direct repeats containing the interacting CGG triplets at both ends. Their DNA-binding module, usually found at the NH2 terminus, comprises a highly conserved zinc binuclear cluster domain, followed by a less conserved linker and then a coiled-coil dimerization motif. AlcR also contains, at its amino terminus, a DNA-binding domain with two zinc atoms linked to six cysteine residues arranged as a compact structure (7, 8). This is similar to that of GAL4, PPR1, and PUT3, whose three-dimensional structures have been resolved (13–16). Mutations in any cysteine of the AlcR zinc cluster result in a loss of DNA binding activity (12). However, AlcR is distinguished by a number of major features which are characterized by *in vitro* binding to single sites (5, 6, 9), containing the consensus core 5'-T/A/GCGG-3' and, *in vivo* by the functional relevance of these sites, organized

---

3 M. Mathieu and B. Felenbok, unpublished results.
DNA Binding Specificity of AlcR Binuclear Cluster  

9801

**FIG. 6. Role of the flanking regions outside the consensus motif for AlcR binding.** A, sequence comparison of AlcR symmetric DNA-binding sites found in various alc gene promoters (Refs. 3-5 and 33, and see Fig. 2). For the sake of consistency, the sequence of the AlcR-binding sites in the *aldA* promoter was inverted with respect to its natural orientation within the promoter region. Positions of the targets with respect to their translation start sites are indicated on the *right*. For simplicity, only the upper strand is shown. Arrows indicate the orientation of the repeated sequence. Conserved nucleotides outside the consensus repeat are boxed. **B,** effect of mutations in the conserved positions outside the inverted repeat target in the *aldA* promoter. Each mutant probe is named after the mutant nucleotide in the corresponding position. Numbers are indicated above the sequence. Mutated nucleotides are marked in boldface type. The complete sequence of the wild type probe is presented under "Experimental Procedures." Relative binding affinity was determined in gel shift experiments as the affinity of the AlcR(1–197) protein for a mutant site compared with that for the single site probe. Binding affinities were estimated as described earlier and compared with that for the single site probe. C, sequence comparison of AlcR DNA-binding sites within direct repeat sites in the *alcA* and *alcR* promoters (see Fig. 3). All the binding sites were orientated in the 5' direction of the TGCGG conserved repeat indicated by an arrow. Site positions relative to their start of translation are indicated on the *right*. Conserved nucleotides flanking the consensus repeat are boxed. **D,** mutational analysis of the flank regions of the artificial AlcR single copy site. Each mutant probe is named after the corresponding mutation. Numbers correspond to the mutated base positions and letters correspond to the保守 bases positions in the consensus sequence. Mutated nucleotides are marked in boldface type. The complete sequence of the probe is present in Table II. The other probes are identical except for the mutated bases (50 fmol) as described under "Experimental Procedures." Binding affinities were relative to AlcR(1–197) binding to the probe.  

in repeated elements, which are defined as UAAS (3, 4, 6, 27).  

One striking structural difference between AlcR and the zinc cluster proteins is the absence of a predicted dimerization domain, consisting of a coiled-coil downstream of the DNA-binding domain (17, 25). This observation allowed us to suggest that, in contrast to most other proteins of this class, AlcR could function as a monomer. In fact, several lines of evidence support the idea that AlcR is a monomer and binds to DNA as a monomer. These include the data presented in this paper as well as *in vitro* and *in vivo* data published previously. First, as follows from our previous *in vitro* studies, the truncated AlcR (1–197) protein which despite containing putative leucine heptads, was unable to form dimers, either in solution or upon binding to DNA (5). According to calibrated gel filtration chromatography the AlcR (1–197) protein has the molecular weight expected for a monomer (results not shown). Furthermore, glutaraldehyde cross-linking data show that the full-length AlcR protein does not form stable dimers in solution. Second, the *in vivo* test with the CI phage repressor clearly indicate that within the entire AlcR sequence no region is able to promote functional dimerization of the CI repressor. Third, *in vitro* binding experiments and analysis of interacting bases also show that a single AlcR molecule is able to bind to a single copy site with the same affinity and the same pattern of base interaction. NMR studies have also shown that one AlcR molecule is bound to a single site (9). The absence of cooperativity upon binding of two AlcR molecules to inverted repeat targets (5, 6) is consistent with AlcR binding DNA as a monomer. Finally, physiological studies by site-directed mutagenesis of AlcR targets on the *alcA* promoter also favor the idea that AlcR acts *in vivo* as a monomer. AlcR was shown to possess a unique specificity, being able to recognize *in vitro* single sites organized as inverted or direct repeats in AlcR controlled promoters. Both types of targets are functional *in vivo* in the *alcA* promoter (6). However, *in vivo* there are no strict requirements for spacing between the direct repeats in the *alcA* promoter and furthermore, the spacing could be increased in such a manner that it would be very difficult to imagine that AlcR binds as a dimer (6). Moreover, disruption of any of the three individual AlcR-binding sites within the target e leaving two other sites intact, lead to an active *alcA* promoter, a result which would not be expected if AlcR is a dimeric protein. Our attempts to directly isolate in *A. nidulans* and in *E. coli* the full-length AlcR protein were unsuccessful. Therefore, we must stress that at this stage only indirect lines of evidence can be presented, which nevertheless when taken together all strongly indicate that AlcR is a monomer.  

The AlcR activator does not possess the linker region present in most proteins of the Cys<sub>2</sub>His<sub>2</sub> class, which directs each protein to its preferred DNA repeated sites depending on the length and composition of the spacer between the half-sites (28). This observation correlates with the finding that spacing between the AlcR repeated binding sites does not seem to affect specific recognition, despite the fact that its length is important in natural palindromic targets. In contrast to most other proteins of this class which bind as dimers, the *in vitro* binding affinity of AlcR was not changed significantly upon alterations of the spacing length. Rather, variations in the number of intervening bases specify the number of AlcR molecules fixed simultaneously on repeated DNA sites. Separated by 2 bp, the interacting cores lie exactly on the opposite sides of the double helix. Such a relative disposition of sites on the DNA makes both of them accessible for contacts with AlcR present in a head to head orientation (see Fig. 5). Placing both sites, whatever their orientation on the same face of the DNA helix, results in binding of only one AlcR molecule to repeated targets. Either the first AlcR molecule when bound to its site sterically blocks the binding of a second molecule or it distorts the DNA, thereby preventing further binding to the adjacent site. Unfortunately, our results do not allow us to discriminate between these hypotheses. In any case, the disappearance of the second complex CII is unlikely explained by the loss of protein-protein interaction. First, increasing the spacing distance restores the forma-
tion of the slowly migrating complex CII containing two AlcR molecules; second, cooperativity of binding has never been observed, as expected if the binding of two AlcR molecules is independent. The total number of AlcR molecules bound to the promoter region might, in turn, be important for synergistic transcriptional activation of the alc genes as has been shown for the aca gene (6).

All the AlcR sites ((T/A)GCGG) in alc promoters encompass the same subsite CGG which is recognized by the other proteins of the zinc cluster family (17, 25). Like most other proteins of this class, AlcR is shown to recognize the same triplet via establishing direct contacts in the major groove of DNA. The zinc cluster proteins all share, between the second and third cysteines, a constant number of highly conserved basic residues. X-ray and NMR studies on GAL4, PPR1, and PUT3 (13–16) have shown that this sequence is the recognition module in which one conserved basic residue at position 4, relative to the third cysteine, makes specific contacts through the N-7 of G residues in the CGG triplet. However, assuming the monomeric structure of AlcR, this triplet alone is obviously not sufficient for binding selectivity. In fact, we show here that bases proximal to the consensus core play an important role in high affinity binding. Thus, the 5′-adjacent G and 3′-A are involved in direct base-specific contacts and, therefore, contribute greatly to binding. This also distinguishes AlcR from GAL4 and PPR1 for which specific nucleotides flanking the recognition triplet are not necessary for binding (30). In the AlcR consensus sequence, the three G residues appeared to be extremely important for binding affinity, which allowed us to extend the recognition unit to the sequence 5′-GCGG-3′. Furthermore, the presence of an AT-rich region 3′ and a purine 5′ to the consensus sequence favors high affinity binding by AlcR. Another regulator, MIG1 which is a Saccharomyces cerevisiae zinc finger protein also requires an adjacent AT-rich region for optimal binding (31). It was speculated that this region facilitates DNA bending and thereby enhances MIG1 affinity for its site as suggested here for AlcR. Another possible explanation for the importance of the flanking regions would be that AlcR establishes specific contacts with DNA outside the core. Although AlcR is indeed able to establish direct contacts, at least with the first 3′ proximal base, the fact that no single base in the 3′-flanking region is essential for binding indicates that a certain DNA structure rather than an unique sequence is important. Our data allow us to propose a new AlcR-binding site: 5′-PuNGCGG-AT rich-3′.

Given that AlcR differs by its structural organization from most other proteins of the same class, one could ask which determinants other than the zinc cluster might define in vivo its binding specificity. One candidate for this role could be the NH2-terminal region adjacent to the zinc cluster. Recent observations have shown both in vitro and in vivo that, unexpectedly outside the AlcR zinc cluster domain, the NH2-terminal region plays a major role in site-specific recognition (32), whereas the downstream region is necessary for high affinity binding (5). An additional, so-called middle homology region may also assist in in vivo binding selectivity. This region was identified by sequence alignment in most of the known zinc cluster proteins and was suggested to participate in DNA target discrimination (17, 26). Preliminary analysis allows us to localize similar stretches of moderate homology within the AlcR sequence. We cannot exclude, however, an alternative explanation involving an accessory protein required for the transcriptional activation of the alc genes.

In conclusion, although it has been proposed previously that some zinc cluster proteins may bind their targets as monomers, AlcR appears to be the first example thus far described, for which detailed evidence, albeit indirect is now available. This gives new insights into the general picture of the zinc binuclear cluster protein family.

Acknowledgments—We thank Prof. B. Holland for the English version of the manuscript, and M. Mathieu for assistance in the construction and test of the rop plasmids.

REFERENCES

1. Felenbok, B., and Sealy-Lewis, H. (1994) Genetics and Physiology of Aspergillus nidulans (Martinelli, S., and Kinghorn, J. R., eds) Vol 29, pp. 141–179, Elsevier Scientific, Amsterdam
2. Fillinger, S., and Felenbok, B. (1996) Mol. Microbiol. 20, 475–488
3. Kulmburg, P., Sequeval, D., Lenouvel, F., Mathieu, M., and Felenbok, B. (1992) Mol. Cell. Biol. 12, 1932–1939
4. Kulmburg, P., Judewicz, N., Mathieu, M., Lenouvel, F., Sequeval, D., and Felenbok, B. (1992) J. Biol. Chem. 267, 21146–21153
5. Lenouvel, F., Nikolaev, I., and Felenbok, B. (1997) J. Biol. Chem. 272, 15521–15526
6. Panazzo, C., Capuano, V., Fillinger, S., and Felenbok, B. (1997) J. Biol. Chem. 272, 22859–22865
7. Kulmburg, P., Prangeé, T., Mathieu, M., Sequeval, D., Scanzocchio, C., and Felenbok, B. (1991) FEBS Lett. 286, 11–16
8. Sequeval, D., and Felenbok, B. (1994) Mol. Gen. Genet. 242, 33–39
9. Cerdan, R., Collin, D., Lenouvel, F., Felenbok, B., and Guittet, E. (1997) FEBS Lett. 408, 235–240
10. Johnston, M., and Dover, J. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 2401–2405
11. Scanzocchio, C. (1994) in Genetics and Physiology of Aspergillus nidulans (Martinelli, S., and Kinghorn, J. R., eds) Vol 29, pp. 259–277, Elsevier Scientific, Amsterdam
12. Acene, I., Lenouvel, F., Sequeval, D., Dexpert, H., and Felenbok, B. (1997) Biochim. Biophys. Acta 1343, 211–220
13. Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S. C. (1992) Nature 356, 408–414
14. Marmorstein, R., and Harrison, S. C. (1994) Genes Dev. 8, 2504–2512
15. Swaminathan, K., Flynn, P., Reece, R. J., and Marmorstein, R. (1997) Nat. Struct. Biol. 4, 751–759
16. Walters, K., Doyle, K. T., Reece, R. J., Ptashne, M., and Wagner, G. (1997) Nat. Struct. Biol. 4, 744–750
17. Scherler, P., and Halmberg, S. (1996) Nucleic Acids Res. 24, 4599–4607
18. Castagnoli, L., Venturi, C., and Cesareni, G. (1994) J. Mol. Biol. 237, 378–387
19. Maxam, A., and Gilbert, W. (1980) Methods Enzymol. 65, 497–559
20. Hu, J. C., O’Shea, E. K., Kim, P. S., and Sauer, R. T. (1990) Science 250, 1400–1403
21. Battaglia, P. A., Longo, F., Ciotta, C., Del Grosso, M. F., Ambrosini, E., and Gigliani, F. (1994) Biochem. Biophys. Res. Commun. 201, 701–708
22. Marchetti, A., Abril-Marti, M., Illi, B., Cesareni, G., and Nasi, S. (1995) J. Mol. Biol. 248, 541–550
23. Strauss, J., Muro Pastor, M. I., and Scanzocchio, C. (1998) Mol. Cell. Biol. 18, 1339–1348
24. Lenouvel, F. (1996) Ph.D. thesis, University of Paris-Sud, France
25. Todd, R. B., and Andrianopoulos, A. (1997) Fungal Genet. Biol. 24, 388–405
26. Porn, O. (1997) Gene (Amst.) 184, 229–235
27. Mathieu, M., and Felenbok, B. (1994) EMBO J. 13, 4022–4027
28. Reece, R. J., and Ptashne, M. (1993) Science 261, 509–511
29. Cerdan, R. (1997) Ph.D. Thesis, University of Paris-Sud, France
30. Liang, S. D., Marmorstein, R., Harrison, S. C., and Ptashne, M. (1996) Mol. Cell. Biol. 16, 3773–3780
31. Lundin, M., Nehlin, J. O., and Ronne, H. (1994) Mol. Cell. Biol. 14, 1979–1985
32. Nikolaev, I., Cochet, M. F., Lenouvel, F., and Felenbok, B. (1999) Mol. Microbiol. 31, 1115–1124
33. Fillinger, S. (1996) Ph.D. thesis, University of Paris-Sud, France

4 I. Nikolaev and B. Felenbok, unpublished results.