The EMBO Journal vol.13 no.2 pp.407-415, 1994

Two different, adjacent and divergent zinc finger binding sites are necessary for CREA-mediated carbon catabolite repression in the proline gene cluster of *Aspergillus nidulans*

Beatriz Cubero and Claudio Scazzocchio[1]

Institut de Génétique et Microbiologie, URA 1354, Université Paris-Sud, Bâtiment 409, 91405 Orsay CEDEX, France

[1]Corresponding author

Communicated by C.Scazzocchio

CREA is the negative regulator mediating carbon catabolism repression in *Aspergillus nidulans*. We have determined all the sites in the DNA region between the *prnD* and *prnB* genes of the proline degradation cluster of this organism which are able to bind a fusion protein containing the zinc finger domain of CREA. The consensus sequence derived for CREA binding is 5′-SYGGRG-3′, but not all possible sites derived from this consensus do in fact bind. The binding of at least some sequences of the form 5′-SYGGAG-3′ is context dependent. Two different and divergent sites, separated by one base pair (5′-GCGGAGACCCCAG-3′), contain the previously sequenced derepressed mutations and are essential for carbon catabolite repression *in vivo*. We have studied the binding of CREA to the region by DNase I and methylation protection, and by methylation and depurination interference. We propose that this pair of sites is the physiological, *cis*-acting element responsible for the carbon catabolite repression of *prnB* transcription.

Key words: *Ascomycetes*/carbon catabolite repression/CREA protein/operator derepressed mutations/zinc fingers

Introduction

The enzymes and the specific permease involved in the utilization of proline as sole carbon and/or nitrogen source are encoded by three genes (*prnD*, coding for proline oxidase, *prnB*, coding for the specific proline permease, and *prnC*, coding for L-Δ^1-pyrroline carboxylate dehydrogenase) clustered in chromosome VII (Arst and MacDonald, 1978; Jones *et al.*, 1981). Also contained in the cluster are the gene coding for the transcription factor mediating specific induction (*prnA*) and a fifth gene of unknown function, *prnX* (Sharma and Arst, 1985; Hull *et al.*, 1989; Gavrias, 1992; see Figure 1a). Transcription of *prnD*, *B* and *C* is strongly dependent on proline induction and requires the integrity of the positively acting *prnA* gene. In addition to this pathway-specific control, the transcription of these genes is also sensitive to carbon catabolite and nitrogen metabolite repression (Sophianopoulou *et al.*, 1993). Repression of transcription is only efficient when both repressing carbon and nitrogen sources are present. Many years ago, derepressed mutations, mapping between the *prnD* and *prnB* genes, were isolated (Arst and Cove, 1973; Arst and MacDonald, 1975). Recently it was shown that these mutations affect the steady state of the transcripts of the *prnB*, *D* and *C* genes (Sophianopoulou *et al.*, 1993). Data from formal genetics studies suggested, but did not prove, that these mutations directly affected the *prnB* gene, and that the simultaneous derepression of the *prnD* and *prnC* genes was a secondary effect resulting from reversal of inducer exclusion (Arst *et al.*, 1980). This problem has not been solved, but at the very least the *prn^d* mutations result in direct derepression of the *prnB* gene. We have sequenced the *prnB* gene, the intergenic region between the divergently transcribed *prnB* and *D* genes and the *prn^d* mutations (Sophianopoulou and Scazzocchio, 1989; Gavrias, 1992; Sophianopoulou *et al.*, 1993). Two mutations, *prn^{d}21* and *prn^{d}22*, are identical; they are both G to A transitions at position −508 (numbering from the *prnB* ATG). The third, *prn^{d}20*, is a G to A transition at position −514 (Sophianopoulou *et al.*, 1993).

It has been postulated, on strictly formal grounds, that the *prn^d* mutations define the binding site for the carbon catabolite repression protein encoded by the *creA* gene (Arst and MacDonald, 1975). The *creA* gene was recently cloned and sequenced (Dowzer and Kelly, 1989, 1991) and shown to code for a protein comprising two zinc fingers which are very similar to those present in the yeast transcriptional regulator, MIG1 (Nehlin and Ronne, 1990). While MIG1 is involved in the carbon catabolite repression of some yeast genes [e.g. *GAL4* and *SUC1* (Nehlin and Ronne, 1990; Nehlin *et al.*, 1991)], CREA is the major and perhaps the universal repressor in *Aspergillus nidulans* (Bailey and Arst, 1975). It was found that in the ethanol regulon of *A.nidulans* a fusion protein containing the *creA* zinc fingers binds *in vitro* to a consensus sequence 5′-SYGGGG-3′, identical to that found for MIG1 (Nehlin *et al.*, 1991; Kulmburg *et al.*, 1993). In this article we show that *prn^{d}22* and *prn^{d}20* define two adjacent and divergent strong CREA binding sites. The *prn^{d}22* mutation maps in a CREA consensus binding site (5′-CTGGGG-3′), while the *prn^{d}20* mutation defines a 5′-GCGGAG-3′ sequence as a physiological binding site. These results extend the CREA and, by implication, the MIG1 binding consensus sequence to 5′-SYGGRG-3′. However, not all the possible sequences derived from the consensus sequence are binding sites. Efficient repression needs the integrity of both the −517 5′-GCGGAG-3′ and the −505 5′-CTGGGG-3′ divergently oriented binding sites.

Results

Mapping the binding sites in the prnB−prnD intergenic region

The intergenic region between the *prnD* and *prnB* genes has been sequenced (Sophianopoulou and Scazzocchio, 1989; Gavrias, 1992). A restriction map of this region is represented by the vertical line in Figure 1c. Figure 1b shows the restriction fragments of the intergenic region which are able to bind the GST::CREA(35−247) fusion protein *in vitro*. Eleven overlapping probes show that there are at least three CREA binding sites in the *prnD−prnB*

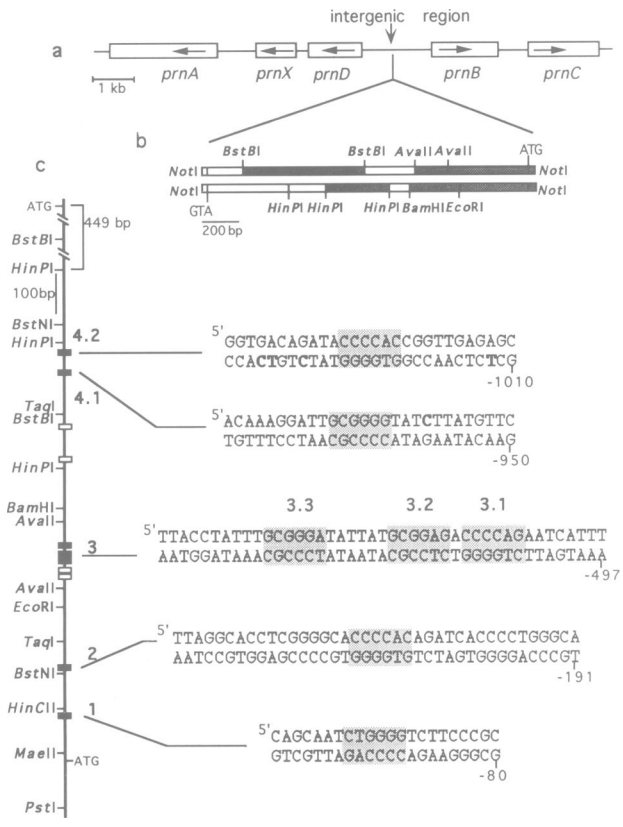

Fig. 1. CREA binding sites in the intergenic region between *prnD* and *prnB*. (**a**) Schematic representation of the proline gene cluster (see text for details). (**b**) Overlapping restriction fragments used in bandshift analysis (results not shown). Open bars, no binding; hatched bars, binding to GST-CREA. ATG codons of *prnD* (left) and *prnB* (right) are indicated. *Not*I sites at the extremities are from the vector polylinker. (**c**) Vertical line, restriction map of the *prnB*–*prnD* region; bottom, *prnB* ATG (Sophianopoulou and Scazzocchio, 1989); top, *prnD* ATG (Gavrias, 1992). Black rectangles, position of protected sites; white rectangles, position of sequences consistent with the consensus, 5′-SYGGRG-3′, but which show no binding. To the right of the vertical line, CREA binding sites, determined by bandshift assays, DNase I (sites 3.1, 3.2, 3.3, 4.1 and 4.2) and methylation protection (all sites), and their flanking sequences. Protection patterns of sites 1 and 2 are shown in Table I, protection patterns of sites 3.1, 3.2 and 3.3 are detailed in Figure 7, and methylation protection patterns of sites 4.1 and 4.2 are shown in Table I. DNase I protection of site 4.1 spans from base −969 to −960 (upper strand) and from base −972 to −955 (lower strand); that in site 4.2 spans from base −1028 to −1019 (lower strand). Bold types, DNase I hypersensitive bands in the protected probes. The restriction fragments used for each of the protection assays were as follows: for site 1, *Bst*NI(−157)–*Pst*I(+135) fragment [α-^{32}P]dCTP-labelled in the non-coding strand, and *Eco*RI(−366)–*Mae*II(−26) [α-^{32}P]dCTP-labelled in the coding strand; for site 2, *Taq*I(−271)–*Hin*cII(−108) [α-^{32}P]dCTP-labelled in the non-coding strand, and *Eco*RI(−366)–*Mae*II(−26) [α-^{32}P]dCTP-labelled in the coding strand; for sites 3.1, 3.2 and 3.3, *Bam*HI(−622)–*Eco*RI(−366) [α-^{32}P]dCTP-labelled in the non-coding strand, and *Ava*II(−592)–*Ava*II(−412) [α-^{32}P]dATP-labelled in the coding strand; for sites 4.1 and 4.2, *Bst*NI(−1106)–*Taq*I(−865) [α-^{32}P]dATP-labelled in the non-coding strand, and [α-^{32}P]dCTP-labelled in the coding strand. The position of each restriction site is given in bp from the adenine of the ATG codon of *prnB*.

intergenic region. Footprinting results show that some probes contain more than one binding site. These experiments show a total of seven binding sites (see below) which have been numbered (from the ATG of *prnB*) 1, 2, 3.1, 3.2, 3.3, 4.1 and 4.2. Symbols carrying the same first digit indicate sites located within the same probe. The *prn^{d}20* and *prn^{d}22* mutations map in the 256 bp *Bam*HI(−622)–*Eco*RI(−366) fragment.

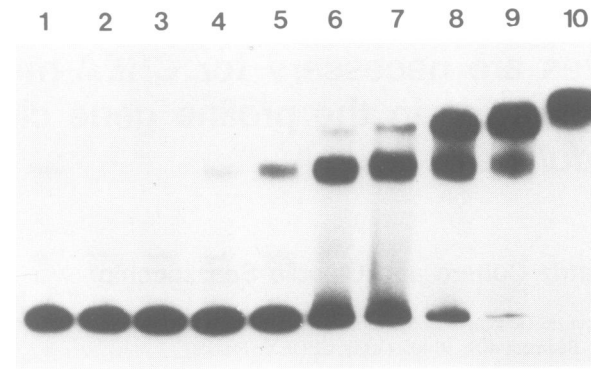

Fig. 2. Gel retardation of the 256 bp *Bam*HI(−622)–*Eco*RI(−366) fragment. Each retardation experiment was conducted with 4 ng of the above fragment labelled as described in Materials and methods and with increasing amounts of CREA fusion protein. Lane 1, no protein; lane 2, 0.3 ng; lane 3, 0.6 ng; lane 4, 3 ng; lane 5, 6 ng; lane 6, 30 ng; lane 7, 60 ng; lane 8, 300 ng; lane 9, 600 ng; lane 10, 3000 ng of protein.

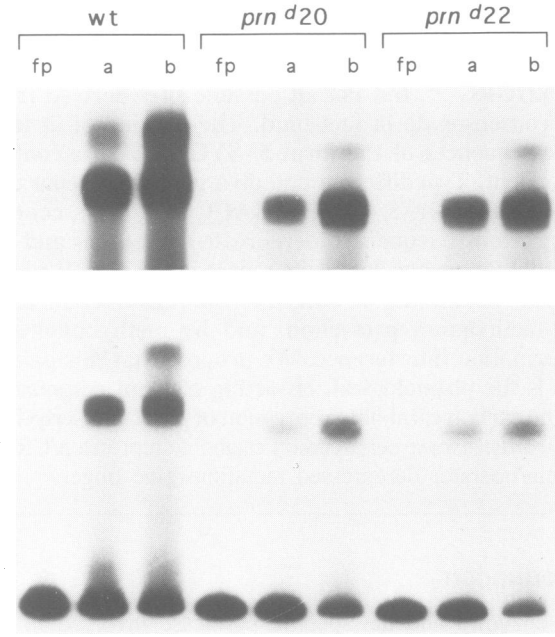

Fig. 3. Gel retardation experiment of wild type and mutant probes. The probe used is the same restriction fragment as in Figure 2. wt, wild type fragment; *prn^{d}20*, equivalent fragment carrying the corresponding mutation; *prn^{d}22*, equivalent fragment carrying the corresponding mutation. In each experiment 4 ng of probe was used. fp, no protein; a, 60 ng of CREA fusion protein; b, 300 ng of fusion protein. The upper panel is an overexposure of the upper part of the autoradiogram in order to reveal the low mobility complex in the probes carrying the *prn^{d}20* and *prn^{d}22* mutations.

Wild type and mutant sites in the 256 bp BamHI(−622)–EcoRI(−366) fragment

Figure 2 shows that this fragment is able to form two complexes of different electrophoretic mobility with the fusion protein. The mutations *prn^{d}20* and *prn^{d}22* do not impair the ability to form either the low or the high molecular weight complexes. However, Figure 3 shows that both the

low and the high molecular weight complexes formed with the mutant probes have slightly higher mobilities than the corresponding complexes formed with the wild type probe. These results can be rationalized if we assume that the

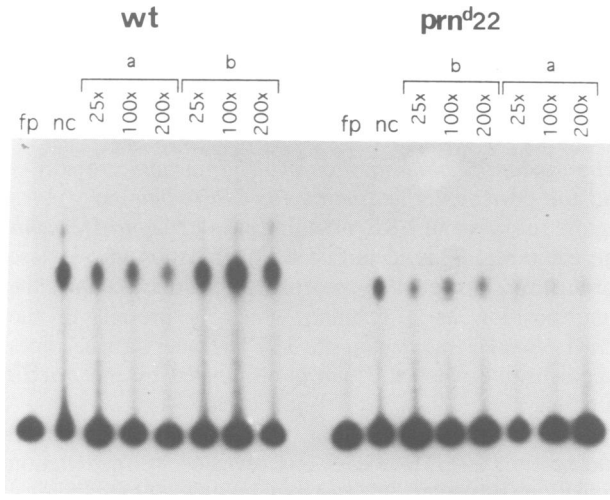

Fig. 4. Competition between the wild type probe and the probe carrying the *prn^{d}22* mutation. In all experiments 4 ng of labelled probe and 60 ng of protein were used. In the left panel (wt) the wild type probe was labelled, in the right panel (*prn^{d}22*) the probe containing the mutation was labelled. fp, free probe, no protein added; nc, non-competed probe, no cold competitor added; panels a, wild type probe used as competitor; panels b, *prn^{d}22* probe used as competitor. The number under the bracket indicates the fold excess relative to the labelled probe.

*Eco*RI−*Bam*HI probe has three binding sites, two of low and similar dissociation constant and a third one of higher dissociation constant, and that each of the mutations results in inability to form one of the 'high affinity' complexes. The footprint experiments reported below show this to be the case.

Figure 4 shows a cross-competition experiment between the wild type probe and one carrying the *prn^{d}22* mutation. In the conditions of this experiment the high mobility complex is the most abundant. This experiment shows qualitatively that the *prn^{d}22* mutation decreases the affinity of the DNA probe for the fusion protein. It should be noted that the difference in mobility between the wild type complex and the mutant complex is considerably less than the one apparent when binding of (presumably) a third molecule to the lower affinity site occurs (see Discussion).

DNase I and methylation protection footprints of the 256 bp *Eco*RI−*Bam*HI fragment show that there are three CREA fusion protein binding sites in this fragment. We called these sites 3.1, 3.2 and 3.3 (see Figure 1c). Mutation *prn^{d}22* lies in site 3.1 and mutation *prn^{d}20* in site 3.2. The experiments shown in Figures 5 and 6 were carried out on the low mobility complexes of the wild type probe and of probes carrying the *prn^{d}22* and *prn^{d}20* mutations respectively. The use of the low mobility complex provides as internal control the binding to site 3.3 (see below) which is not protected in the high mobility complex, while sites 3.1 and 3.2 are protected (results not shown). The results are summarized in Figure 7. The *prn^{d}22* mutation results in complete loss of binding to site 3.1 and *prn^{d}20* in complete loss of

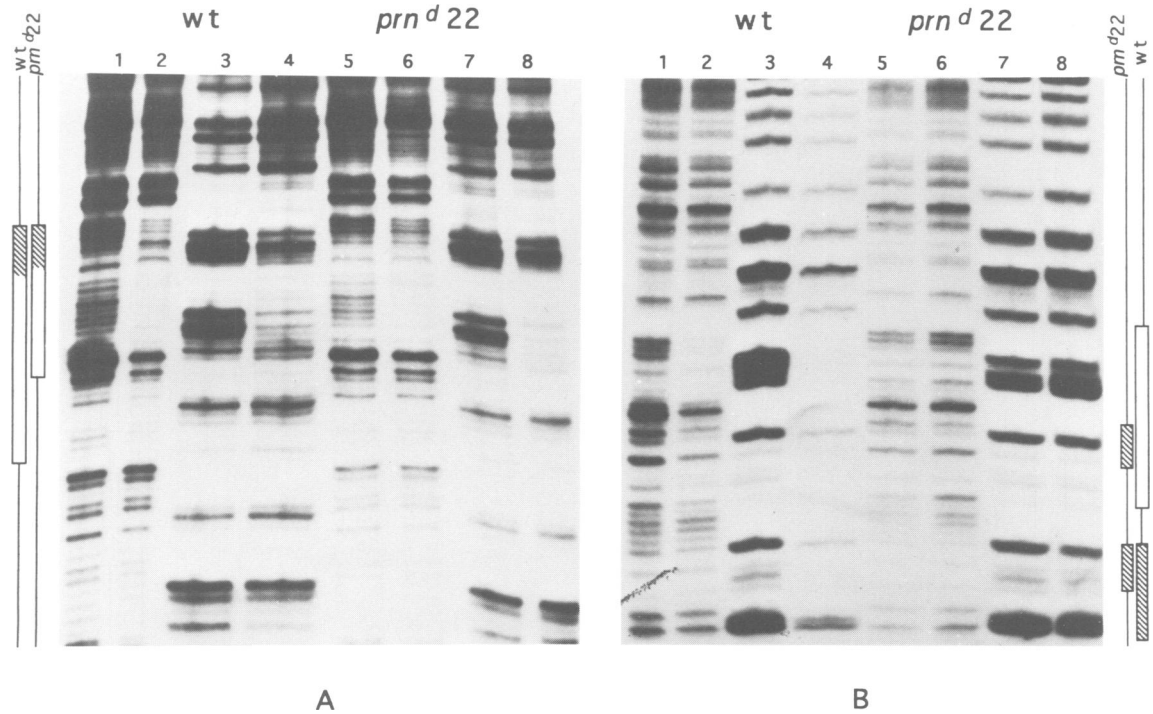

Fig. 5. Footprints of the wild type and of the probe carrying the *prn^{d}22* mutation. (**A**) Non-coding strand; (**B**) coding strand [coding and non-coding refer to the *prnB* gene (Sophianopoulou and Scazzocchio, 1989)]. Lanes 1−4, wild type probe; lanes 5−8, probe carrying the *prn^{d}22* mutation. Lanes 1 and 5, DNase I digestion, non-protected probe; lanes 2 and 6, DNase I digestion, protected probe; lanes 3 and 7, methylation with DMS, non-protected probe; lanes 4 and 8, methylation with DMS, protected probe. In panel (B) the complete loss of binding to site 3.1 is evident; the loss of one G in the probe carrying the mutation *prn^{d}22*, which is a G:C to A:T transition (Sophianopoulou *et al.*, 1993), is also clear. At the side of each panel, open boxes represent strongly protected regions, while hatched boxes or areas indicate partially protected regions, which correspond exactly to the brackets in Figure 7.

binding to site 3.2. It is noteworthy that sites 3.2 and 3.3 show protection of the G in the second position in the C-rich strand of the site. Interestingly, the *prn^{d}22* mutation also affects the methylation protection pattern of the Gs in the C-rich strand of sites 3.2 and 3.3 and each mutation affects the DNase protection pattern of the C-rich strand of the other two sites (see Discussion). The appearance of DNase I and methylation hypersensitive bands in the mutant *prn^{d}20* may be of significance; indeed, the mutant T is hypersensitive in the DNase I protection experiment. This may reflect changes in DNA conformation due to the occupation of the other two CREA binding sites.

We investigated the role of the purine bases in both strands in this region by missing-base depurination and by methylation interference. These techniques gave completely consistent results, shown in Figure 8. In both sites 3.1 and 3.3 the four Gs of the G-rich strand are necessary for CREA binding. For sites 3.2 and 3.3 we found that both methylation and loss of the three Gs and, strikingly, of the G in the opposite strand (complementary to the C in the second position) interfere partially (~50% by densitometric scan analysis) with CREA binding. The loss of the A in site 3.2 (5′-GCGGAG-3′) results in a partial (~50% by densitometric scan analysis) interference with binding. In site 3.3 we found that the A in position 6 does not participate in CREA binding.

Surprisingly, we found that loss of the two As 5′ to the G-rich strand of site 3.2 interferes with CREA binding to the same extent as the A inside the site. All partially interfering Gs and As were evident in a number of independent experiments. Quantitative densitometry (see Materials and methods) was carried out in two independent experiments with completely consistent results.

Other potential binding sites in the intergenic region and the consensus sequence for CREA binding

Figure 1c shows all CREA binding sites in the *prnD−prnB* intergenic region as identified by methylation and DNase I protection. In site 1 a hypersensitive guanine occupies the same position as a guanine which is present in the *Eco*RI−*Bam*HI fragment in site 3.2. The latter base becomes hypersensitive when this site is eliminated by the *prn^{d}20* mutation.

In fragment *Taq*I(−271)−*Hinc*II(−108) a canonical binding site, 5′-GTGGGG-3′, is revealed by methylation protection. However, in some experiments a non-canonical 8 bp site on the other strand, 5′-CTCGGGGC-3′, also shows weak protection.

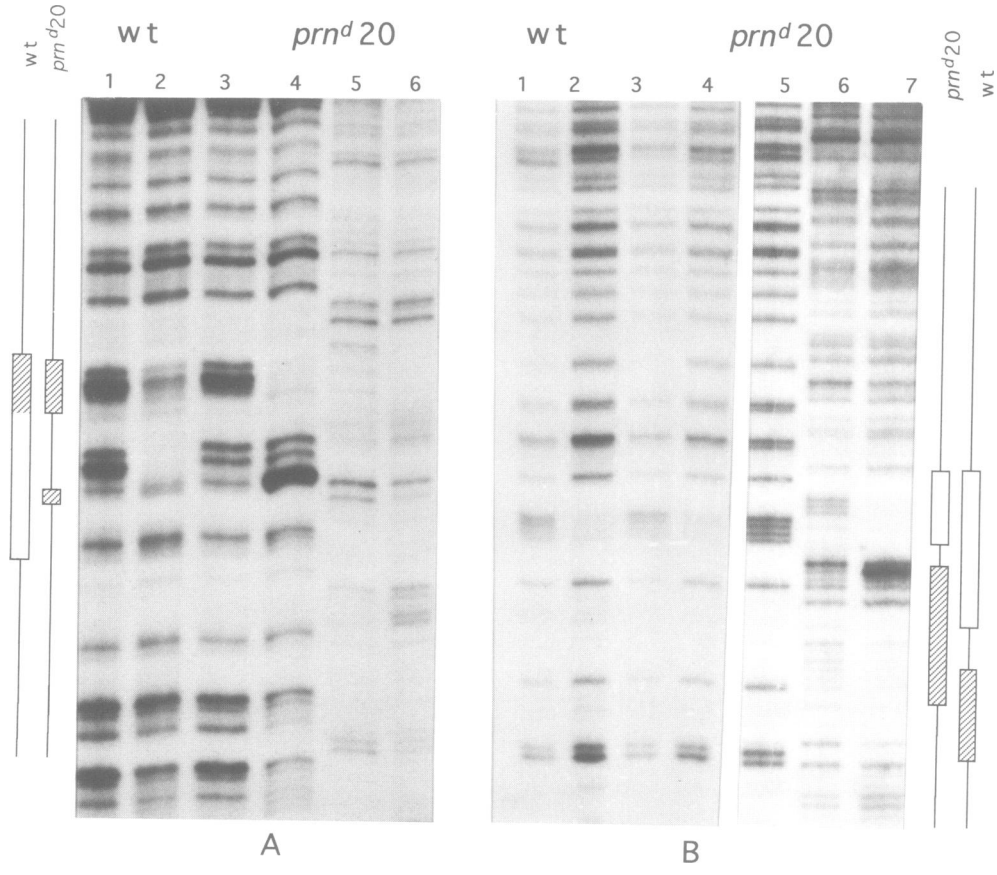

Fig. 6. Footprints of the wild type and of the probe carrying the *prn^{d}20* mutation. **(A)** Non-coding strand; **(B)** coding strand [coding and non-coding refer to the *prnB* gene (Sophianopoulou and Scazzocchio, 1989)]. Panel A: lanes 1 and 3, methylation with DMS, non-protected probe; lanes 2 and 4, methylation with DMS, protected probe; lanes 1 and 2, wild type probe; lanes 3 and 4, probe carrying the *prn^{d}20* mutation; lanes 5 and 6, DNase I digestion of the probe carrying the mutation *prn^{d}20* (lane 5, non-protected probe; lane 6, protected probe). The wild type probe shown in Figure 5 was not repeated in this experiment. Panel B: lanes 1−4 as in panel A; lane 5 is the same as lane 3 and is included as a guanine ladder for the DNase I experiment; lanes 6 and 7, DNase I protection of the probe carrying the *prn^{d}*20 mutation (lane 6, free probe; lane 7, bound probe). The band which is strongly hypersensitive to methylation in lane 4 is the sixth guanine of site 3.2. The DNase I hypersensitive band in lane 7 is the T corresponding to the prn^{d}*20* mutation in site 3.2. At the side of each panel, open boxes show strongly protected regions, and hatched boxes or areas show partially protected regions, which correspond exactly to the brackets in Figure 7.

The binding consensus for the CREA protein has been established by DNase I protection of probes derived from the *A.nidulans* ethanol regulon as 5′-SYGGGG-3′. The results of this article extend this consensus sequence to 5′-SYGGRG-3′. However, both in the ethanol regulon and in the proline cluster some sites agreeing with this consensus show no binding to the CREA fusion protein. Table I summarizes all the potential CREA binding sites for the two *A.nidulans* systems. From this table it would follow that a C rather than a T at position 2 is essential for CREA binding when an A rather than a G is present in position 5. In order to substantiate this conclusion we have constructed the oligomers which place both a 5′-GTGGAG-3′ and 5′-CCGGAG-3′ sequence in the same sequence context as site 3.2 in a *prn^{d}22* background, which eliminates binding to the immediately adjacent site, 3.1. These results are shown in Figure 9 (left). The 5′-GTGGAG-3′ sequence binds CREA, surprisingly even better than site 3.2, when placed in this sequence context, while the 5′-CCGGAG-3′ sequence shows no binding even in this context. Figure 9 (right) shows the sequences upstream of all the CREA binding sites found in the *prn* cluster and in genes *alcA* and *alcR*, also regulated by CREA (Kulmburg *et al.*, 1993). No obvious consensus flanking sequences have been found for 5′-SYGGGG-3′ sites.

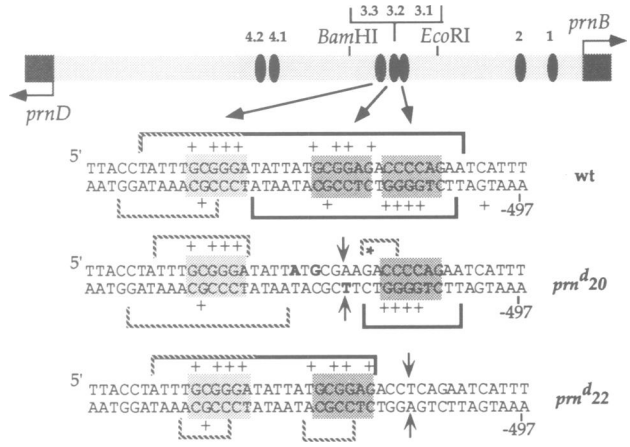

Fig. 7. Top: schematic representation of the *prnD−prnB* intergenic region. The position of all CREA binding sites is indicated by grey ovals. *Bam*HI and *Eco*RI restriction sites are indicated. **Bottom**: protection pattern of the wild type *Bam*HI(−622)−*Eco*RI(−366) probe and of probes carrying the *prn^{d}20* and *prn^{d}22* mutation respectively. The base pairs altered in these mutations are indicated by arrows. Shaded boxes, CREA binding sites; continuous brackets, strong DNase I protection; dotted brackets, weak DNase I protection; bold type, hypersensitivity to DNase I; (+), guanines protected from methylation; (*), guanines hypersensitive to methylation in the bound DNA.

Discussion

We have determined by DNase I footprinting, methylation protection and, for the 3.1, 3.2 and 3.3 sites, depurination and methylation interference, the binding sites for a CREA fusion protein in the intergenic region of the divergently transcribed proline cluster of *A.nidulans*. Two divergently oriented sequences, separated by one base pair, are necessary for carbon catabolite repression to occur.

The present data establish the CREA binding consensus as 5′-SYGGRG-3′. The MIG1 protein also binds to a 5′-GCGGAG-3′ sequence albeit less tightly than to a 5′-GCGGGG-3′ sequence (M.Lundin, J.O.Nehlin and H.Ronne, personal communication). The importance of a G at position 6 is confirmed by the properties of site 3.3 (5′-GCGGGA-3′). This site is not protected in the high mobility complex, is only partly protected in the low mobility complex and forms at a concentration of protein ~10 times higher than the high mobility complex (Figure 2).

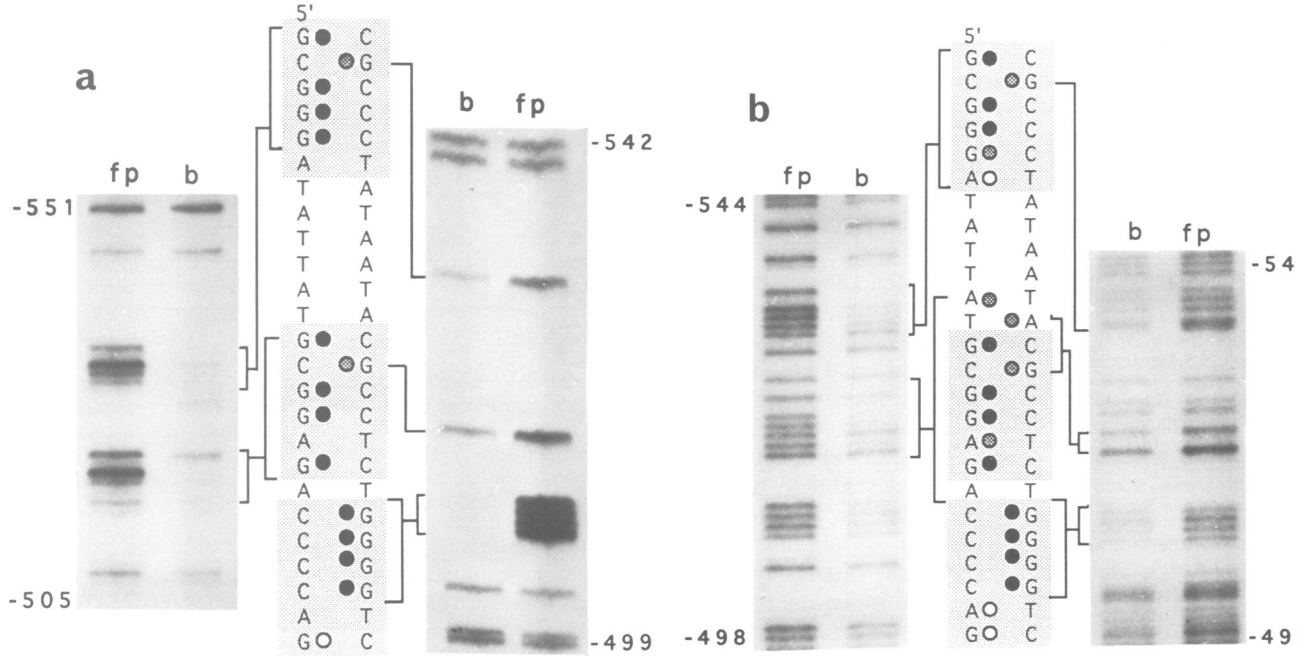

Fig. 8. Methylation (**a**) and depurination (**b**) interference footprints of the DNA fragment which contains sites 3.1, 3.2 and 3.3. For the sake of clarity only the region corresponding to the three CREA binding sites is shown (no interference was found elsewhere). Left panel, coding strand; right panel, non-coding strand. Black dots, strongly interfering guanines (a) or purines (b); hatched dots, partially interfering guanines (a) or purines (b); white dots, guanines or purines which in principle form part of the 6 bp CREA binding sequence but which show no interference. The three binding sites are highlighted by shaded boxes. fp, free probe, no protein added; b, bound probe, 100 ng of protein added.

Table I. Binding ability of all possible CREA binding sites consistent with the consensus sequence 5′-SYGGRG-3′

Binding site sequence	*prn* intergenic region		*alcR*		*alcA*	
	Pres.	Prot.	Pres.	Prot.	Pres.	Prot.
5′ **G C G G G G** / **C G C C C C**	+	+	+	+	+	+
G T G G G G / C A C C C C	+	+	−		+	nd
C C G G G G[a] / G G C C C C	−		+	nd	−	
C T G G G G / **G A C C C C**	+	+	+	+	+	nd
G C G G A G / **C G C C T C**	+	+	+	+	+	nd
G T G G A G / C A C C T C	+	−	+	nd	+	nd
C C G G A G / G G C C T C	+	−	+	−	−	
C T G G A G / G A C C T C	+	−	+	−	−	

(+) indicates presence (Pres. column) of the site in the promoter or protection (Prot. column) against DNase I (*prn* intergenic region and *alc* genes) or DMS methylation (*prn* intergenic region). Protected guanines are indicated by stars. Absence of the site (Pres. column) or no binding (Prot. column) is indicated by (−). nd, not determined. Sequences that have been shown to be functional *in vivo* are in bold. The low affinity site 3.3 (5′-GCGGGA-3′) has not been included in this table. *alcA* and *alcR* data are from Kulmburg *et al.* (1993).

[a]Espeso and Peñalva have found this site upstream of the *ipnsA* (isopenicillin synthase) gene of *A.nidulans*. This site shows no binding to CREA (E.Espeso and M.A.Peñalva, personal communication). Thus, all the sites consistent with the 5′-SYGGRG-3′ consensus have been investigated.

Depurination experiments show that this adenine is not involved in CREA binding.

CREA belongs to the group of zinc finger proteins exemplified by Zif268 (Dowzer and Kelly, 1991). If we assume that the binding of the two CREA zinc fingers proceeds in a pattern similar to the binding of the three zinc fingers of Zif268 (Pavletich and Pabo, 1991) it is straightforward to account for CREA binding to the sequence 5′-GYGGGG-3′. Methylation protection results of such sites (2, 4.1 and 4.2) are consistent with this binding pattern (Figure 1 and Table I). The crystallographic studies of Pavletich and Pabo (1991) have shown that each zinc finger recognizes a module of 3 bp. Jacobs (1992) has generalized the model of Pavletich and Pabo (1991) and observed that usually two of three positions [called s3, s6 and m3 (Jacobs, 1992)] in each zinc finger of many different proteins are occupied by residues that can interact directly with DNA base pairs. The amino-terminal zinc finger of CREA has residues that could form hydrogen bonds with guanines at each of these three positions. Positions s3 and m3 are arginines and position s6 is a histidine. In the carboxy-terminal finger, positions s3 and m3 are arginines, and position s6 is a glutamic acid. Thus the binding residues would be RHRR-R which would bind in an 'antiparallel' fashion (the most amino-terminal residue binding the most 3′ guanine) to 5′-GXGGGG-3′. The proposed pattern of the binding of CREA zinc fingers to sequences of the 5′-GYGGGG-3′ type is shown in Figure 10. The binding and the protection pattern to 5′-CYGGGG-3′ sites could be rationalized by the same model if we assume that the carboxy-terminal finger needs to bind only one base when the amino-terminal finger is binding three. Binding to only a conserved core GGA seems to suffice for the two-fingered Tramtrack protein of *Drosophila melanogaster* (Fairall *et al.*, 1992). Mutational analysis has shown that the middle zinc finger of Krox-20, a protein belonging to the Zif268 family [critical residues: s3R, s6H and m3T; nomenclature as above (Jacobs, 1992)], can bind to both 5′-GGG-3′ and 5′-GAG-3′ (Nardelli *et al.*, 1992).

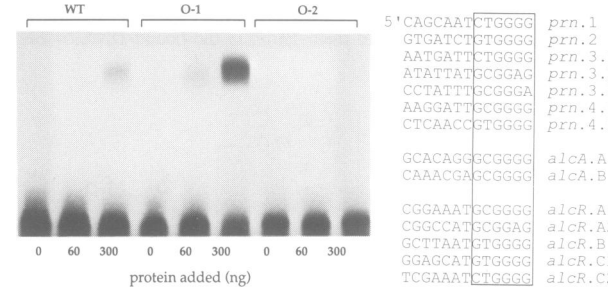

Fig. 9. Left: gel retardation experiments with double-stranded 22mer oligonucleotides carrying different 5′-SYGGAG-3′ sites in the context of binding site 3.2 (5′-GCGGAG-3′) in a *prn*^{d}22 background. The sequences of the oligonucleotides (coding strand) are: wt, 5′-GATATTATGCGGAGACCTCAGA-3′; O-1, 5′-GATATTATGTGGAGACCTCAGA-3′; O-2, 5′-GATATTATCCGGAGACCTCAGA-3′. The 6 bp consensus-derived sequences are underlined. See Table I for binding properties of these sites in their native context. **Right**: comparison of the upstream sequences of all CREA binding sites (see Discussion) in the *prn* intergenic region (this work) and the *alc* regulon of *A.nidulans* (Kulmburg *et al.*, 1993), as determined by bandshift assays, DNase I protection (all sites) and methylation protection (*prn* sites).

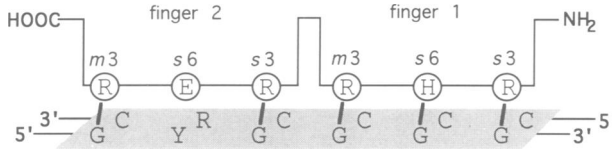

Fig. 10. Proposed binding pattern of the CREA zinc finger protein to a binding site of the 5′-GYGGGG-3′ class. The relevant amino acids are defined by the notation of Jacobs (1992).

The binding of CREA to the 5′-SYGGAG-3′ sites deserves some discussion. In principle the histidine at position s6 of the first finger could share a proton with the N7 of either an adenine or a guanine residue but also with the O6 of this base (Pavletich and Pabo 1992). The partial interference obtained by removing the adenine (compared with the adjacent guanines) in the 5′-GCGGAG-3′ 3.2 site suggests that, while histidine s6 may well bind adenine, this bond may not be essential for the formation of the CREA−DNA complex. A comparison of the 5′ flanking sequences of all the sites that bind CREA *in vitro* [this work for the proline cluster; Kulmburg *et al.* (1993) for the *alc* regulon] strongly suggests that, whereas the binding ability of 5′-SYGGGG-3′ sequences is absolute and context independent and the same can be said of sequence 5′-GCGGAG-3′, other 5′-SYGGRG-3′ sequences only bind CREA when placed in an AT-rich 5′ context and when the base at position 1 is a guanine (Table I and Figure 9).

A similar dependence on flanking sequences has been found for MIG1 by M.Lundin, J.O.Nehlin and H.Ronne (personal communication) using synthetic oligonucleotides. A confirmation of the role of flanking sequences is given by the results of the depurination experiments, where the loss of two As external to the consensus sequence interfere, albeit partially, with CREA binding.

It is noteworthy that both in the high affinity site 3.2 and in the low affinity site 3.3 a strong protection and a partial interference is found for the G at position 2 (Figure 7), a G which is in the C-rich strand. There seems to be a correlation between the loss of a contact in the 3′ base triplet and the establishment of a new contact in the opposite strand of the 5′ base triplet. Moreover, when a G in position 2 in the C-rich strand interferes, the G in position 1 interferes very weakly. This may suggest a distortion of the α-helix of the finger. It also suggests that the fingers may not be acting as completely independent modules. This is also suggested by Pavletich and Pabo (1991). However, Figure 7 also shows that when the binding to the 3.1 site is lost, the protection of Gs in the C-rich strand is strongly diminished (virtually lost for the G of site 3.2; still visible in lower exposures of the gel for the G of site 3.3). This may suggest that the contact with the G in the C-rich strand, rather than being an intrinsic property of sites with a C in position 2, could be an indirect effect of the presence of an adjacent strong divergent binding site. The model of Pavletich and Pabo (1991) states that the tandemly arranged zinc fingers should contact bases in the same strand. We note that Lemaire *et al.* (1990) have reported a weak interference of a G in position 8 (or 9 or 10, see below) of the C-rich strand of a G-rich probe bound to the three-fingered Krox24, Krox20 (Zif268) and Sp1 proteins. The interference is partial but clear, but no equivalent contact is seen in the X-ray diffraction studies of Pavletich and Pabo (1991). It must be pointed out that the probes used in these two studies did not have the same sequence and that the probe used by Lemaire *et al.* (1990) contained at least three overlapping binding sites comprising 11 bp. It remains to be established if the contacts in the C-rich strand indicate that sites of the form 5′-SYGGGG-3′ bind strictly according to the model of Pavletich and Pabo (1991), while those of the form 5′-GYGGAG-3′ do not, or whether the apparent contact in the C-rich strand is due to changes in the DNA structure resulting from the proximity or overlap of several binding sites. That the occupancy of two adjacent sites results in changes in DNA structure is strongly suggested by the pattern of methylation hypersensitivity. There is a hypersensitive guanine at position −87 outside site 1. There is a guanine in the same relative position outside site 3.1. This guanine is not hypersensitive when both sites 3.1 and 3.2 are occupied, but it becomes hypersensitive when only site 3.1 is occupied in the *prn^{d}20* mutant.

It is also noteworthy that a two-finger structure in which the amino-terminal finger has base-pair binding amino acids at three positions (s3, s6 and m3) and the other finger has base-pair binding amino acids at two positions (s3 and m3) seems to be characteristic of the Ascomycetes. As well as in CREA and MIG1, such a structure is found in the yeast proteins ADR1 and RGM1, and in the *A.nidulans* BRLA (Hartshorne *et al.*, 1986; Estruch, 1991; Adams *et al.*, 1988). The three-fingered mammalian genes of the Zif268 family have only two positions in each finger occupied by potential base-pair binding residues, usually in the order s3R,m3R, s3R,s6H, s3R,m3R. In fact the s3R,s6H,m3R structure is absolutely exceptional in higher eukaryotes [e.g. the fourth finger of the 12-fingered 'Suppressor of hairy wing' in *D.melanogaster* (Parkhurst *et al.*, 1988) and the fourth finger of the 13-fingered human MZF1 protein (Hromas *et al.*, 1991)].

The fact that in the wild type probe simultaneous binding occurs at the two divergent sites 3.1 and 3.2 suggests that cooperative or simultaneous binding might occur. This may be due to the fact that glutathione thiotransferase is normally a dimer (Mannervik, 1985), thus we may have a fusion protein dimer which may recognize with high affinity a pair of divergent, adjacent sites. This could neatly explain why there is a surprisingly small difference in mobility between the wild type high mobility complex, where the two high affinity sites are occupied, and the mutant complex, where only one site is occupied. In both cases the DNA probe would be bound by a dimeric fusion protein, and the small difference observed would be due not to the number of protein molecules bound but to a conformational difference between a dimer that bound two sites in DNA and a dimer that bound only one. This presumed dimerization artefact may not be totally irrelevant to the situation *in vivo*, where dimerization of the native CREA protein may also occur. While CREA, like all other zinc finger proteins of the Cys-Cys/His-His type, is able to bind *in vitro* to single sites, it would recognize as the physiological sequence upstream of *prnB* a pair of divergent sites, exactly as has been shown for the ADR1 binding sites upstream of the ADHII gene of *S.cerevisiae* (Yu *et al.*, 1989). The results of the footprints strongly suggest that the binding to the two divergent sites is not independent. Not only is the protection of the G in position 2 of site 3.2 lost when the binding to site 3.1 is lost (see above), but each mutation results in drastic

modification of the DNase I protection for the C-rich strand of the adjacent site (Figure 7).

The most significant result of this work is that in the *prnD–prnB* intergenic region there are two different and divergent CREA binding sites (sites 3.1 and 3.2). The mutations defined as operator derepressed by early genetical (Arst and MacDonald, 1975; Arst *et al.*, 1980) and recent molecular data map precisely in these sites (Sophianopoulou *et al.*, 1993). Site 3.1 is identical to site 1 upstream of GAL4 and has been shown by a disruption experiment to be necessary for MIG1 mediated repression of GAL4 (Griggs and Johnston, 1991). It may be noteworthy that all the sites that have been shown to be functional *in vivo* are preceded by AT-rich stretches (Griggs and Johnston, 1991; Nehlin *et al.*, 1991; this work). In the GAL4 promoter there is a second MIG1 site 9 bp downstream from the 5′-CTGGGG-3′ sequence. This site, 5′-CTGGAG-3′, is preceded by an AT-rich sequence (TGTTA; compare with ATTAT in the binding oligomer of Figure 8). Deletion of this sequence results in partial derepression of GAL4 and an increase of the distance between the two sites by the insertion of 6 bp results in increased repression (Griggs and Johnston, 1991).

The single transitions *prn^{d}20* and *prn^{d}22* completely eliminate binding to each of the divergent sites. These two sites (3.1 and 3.2; Figure 7) encompass the sequence described by Sophianopoulou *et al.* (1993; 5′-GAG-ACCCC-3′ or its complement, 5′-GGGGTCTC-3′) as the minimal repression site. This pair of divergent sites is necessary for the efficient *in vivo* repression of the *prnB* gene and directly or indirectly for the repression of the *prnD* and *prnC* genes. There is strong evidence that it is also sufficient. Loss of binding at site 3.2 results in partial derepression, while loss of binding at site 3.1 results in complete derepression (Sophianopoulou *et al.*, 1993) Preliminary deletion experiments encompassing both sites result in strikingly higher than wild type and complete glucose-derepressed expression of the *prnB* gene (V.Gavrias and C.Scazzocchio, unpublished results). The presence of the additional CREA binding sites has been useful for us to determine the minimal requirements for CREA binding. The agreement between the early genetical work and the footprinting analysis is satisfying and validates the results obtained *in vitro* with the CREA fusion protein and, by extension, the MIG1 protein by us and other workers (Nehlin and Ronne, 1990; Nehlin *et al.*, 1991; Kulmburg *et al.*, 1993).

Materials and methods

Strains and plasmids

Escherichia coli, strain BL21 [*hsdS gal*(λcIts857 *ind*1 Sam7 *nin*5 *lac*UV5-T7 gene *1*)] was used for the expression of GST-CREA fusion protein, and the standard strain JM109 for plasmid propagation. Probes from the right part of the intergenic region were obtained by cloning a 759 bp *Bam*HI–*Pst*I fragment of plasmid pAN222 (Hull *et al*, 1989), which goes from position −619 upstream of the gene *prnB* to position +136 inside its open reading frame, into the *Sma*I site of Bluescript KS+. The same restriction fragments of the corresponding λEMBL4 clones (Sophianopoulou *et al.*, 1993) were used for subcloning mutations *prn^d*20 and *prn^d*22. Probes from the left part of the intergenic region of the *prn* cluster were obtained from plasmid pAN224 (Hull *et al.*, 1989) with the appropriate restriction enzymes. All DNA manipulations were performed following standard procedures (Sambrook *et al.*, 1989).

Expression of GST::CREA(35–247) fusion protein

E.coli BL21 was transformed with the pGEX-*creA* plasmid, which carries the 619 bp *Nco*I–*Nco*I fragment containing the zinc finger domain of CREA in the expression vector pGEX2T (Kulmburg *et al.*, 1993). The expression and purification of the GST-CREA fusion protein was performed following the method described by Smith and Johnson (1988).

Gel retardation assays

DNA restriction fragments were end-labelled with the appropriate [α-^{32}P]dNTP using Sequenase version 2.0 (USB, Cleveland, OH) and purified from a non-denaturing 6% polyacrylamide gel. Double-stranded, [α-^{32}P]dNTP-labelled, 22mer oligonucleotides were prepared by annealing the 22mer coding strand (coding refers to *prnB*) to a corresponding 15mer complementary strand and subsequently filling in the protruding 5′ ends using Klenow enzyme. Binding assays were performed in 20 μl reaction mixtures containing 4 ng of labelled DNA, 2 μg of poly(dI-dC), 25 mM Tris–HCl (pH 8.0), 100 mM KCl, 4 mM spermidine, 10% glycerol and 60 ng of gluthatione–Sepharose purified GST-CREA protein. After incubation at 25°C for 15 min the reaction mixtures were run at 18 V/cm through a non-denaturing 6% polyacrylamide–10% glycerol gel in 0.25 × TBE buffer at 4°C. The gels were transferred to Whatman 3MM paper, dried and subjected to autoradiography. For competition experiments different amounts of unlabelled DNA were added to the reaction mixtures prior to addition of the protein.

DNase I footprinting

Binding reactions were performed as described above and incubated with 2.5 mM $CaCl_2$, 5 mM $MgCl_2$ and 2 ng of DNase I (Boehringer Mannheim) for 1 min at 25°C. Digestions were stopped with 0.5 μmol of EDTA (pH 8.0) and the samples immediately purified by acrylamide gel electrophoresis as in gel retardation assays. The gels were autoradiographed and the bound and free probes were excised from the gel and electroeluted onto DEAE membranes (Schleicher and Schuell, Dassel). After phenol and chloroform extractions the digested DNA was analysed in a 6% polyacrylamide–urea sequencing gel. Nucleotide positions were identified by the Maxam and Gilbert reaction (Maxam and Gilbert, 1980) for guanines run in parallel. Protected and hypersensitive bases were identified by densitometric scanning of autoradiograms.

Methylation protection footprinting

Binding reactions were performed as described above, and DNA was partially methylated by subsequent incubation with 0.25% DMS in cacodylate buffer (50 mM sodium cacodylate pH 7.0, 1 mM EDTA pH 8.0) at 25°C for 5 min; free and bound DNA were separated and recovered as described above and cleaved by piperidine at the G residues (Maxam and Gilbert, 1980). The reaction products were analysed on a 6% polyacrylamide–urea sequencing gel.

Methylation interference assays

End-labelled [α-^{32}P]dNTP probes were partially methylated by incubation with 0.25% DMS in cacodylate buffer (see above) and 2 μg poly(dI-dC) at 25°C for 5 min. Binding reactions were performed as described above and free and bound DNA were separated, recovered and cleaved by piperidine at the G residues (Maxam and Gilbert, 1980). The reaction products were analysed on a 6% polyacrylamide–urea sequencing gel. Differences in labelling of the lanes were normalized by equating the absorbance of bands clearly outside the DNase I protected regions for the free and bound lanes.

Depurination interference assays

Depurination mixtures contain 10 ng of labelled probe, 1 μg of yeast tRNA and 0.2 N formic acid in a volume of 10 μl. After treatment for 1 min at 65°C reactions were stopped with 30 μl of 3 M ammonium acetate pH 7.5. DNA was ethanol precipitated, recovered by centrifugation, resuspended in 20 μl of water, precipitated twice and recovered with 30 μl of 1.5 M sodium acetate pH 5.3 and 2 vols of ethanol. Binding reactions and gel retardation assays were performed as described above; bound and free probes were recovered from the gel by electroelution and cleaved with piperidine. The reaction products were separated on a 6% polyacrylamide–urea sequencing gel and densitometric analysis was performed as above.

Acknowledgements

We thank V.Gavrias for the unpublished sequence between the *Hin*dIII site at −699 of *prnB* and the ATG of *prnD*, P.Kulmburg for the gift of plasmid pGEX-*creA*, B.Felenbok for helpful discussions and J.Strauss for invaluable technical advice and helpful discussions. We thank H.Ronne and M.A.Peñalva for unpublished data. This work was supported by the Université Paris-Sud, the CNRS and by a Bridge contract (BIOT CT90-0169) from the European Community. B.Cubero was supported by a postdoctoral fellowship from the Spanish Ministerio de Educacion y Ciencia.

References

Adams,T.H., Boylan,M.T. and Timberlake,W.E. (1988) *Cell*, **54**, 353–362.

Arst,H.N.,Jr and Cove,D.J. (1973) *Mol. Gen. Genet.*, **126**, 111–141.

Arst,H.N.,Jr and MacDonald,D.W. (1975) *Nature*, **254**, 26–31.

Arst,H.N.,Jr and MacDonald,D.W. (1978) *Mol. Gen. Genet.*, **163**, 17–22.

Arst,H.N.,Jr, MacDonald,D.W. and Jones,S.A. (1980) *J. Gen. Microbiol.*, **116**, 285–294.

Bailey,C. and Arst,H.N.,Jr (1975) *Eur. J. Biochem.*, **51**, 573–577.

Dowzer,C.E.A. and Kelly,J.M. (1989) *Curr. Genet.*, **15**, 457–459.

Dowzer,C.E.A. and Kelly,J.M. (1991) *Mol. Cell. Biol.*, **11**, 5701–5709.

Estruch,F. (1991) *Nucleic Acids Res.*, **19**, 4873–4877.

Fairall,L., Harrison,S.D., Travers,A.A. and Rhodes,D. (1992) *J. Mol. Biol.*, **226**, 349–366.

Gavrias,V. (1992) Ph.D. thesis, Université Paris-Sud, Orsay, France.

Griggs,D.W. and Johnston,M. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 8597–8601.

Hartshorne,T.A., Blumberg,H. and Young,E.T. (1986) *Nature*, **320**, 283–287.

Hromas,R., Collins,S.J., Hickstein,D., Raskind,W., Deaven,L.L., O'Hara,P., Hagen,F.S. and Kaushansky,K. (1991) *J. Biol. Chem.*, **266**, 14183–14187.

Hull,E.P., Green,P.M., Arst,H.N.,Jr and Scazzocchio,C. (1989) *Mol. Microbiol.*, **3**, 553–559.

Jacobs,G.H. (1992) *EMBO J.*, **11**, 4507–4517.

Jones,S.A., Arst,H.N.,Jr and MacDonald,D.W. (1981) *Curr. Genet.*, **3**, 49–56.

Kulmburg,P., Mathieu,M., Dowzer,C., Kelly,J. and Felenbok,B. (1993) *Mol. Microbiol.*, **7**, 847–857.

Lemaire,P., Vesque,C., Schmitt,J., Stunnenberg,R.F. and Charnay,P. (1990) *Mol. Cell. Biol.*, **10**, 3456–3467.

Mannervik,B. (1985) *Adv. Enzymol.*, **57**, 357–417.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol*, **65**, 499–560.

Nardelli,J., Gibson,T. and Charnay,P. (1992) *Nucleic Acids Res.*, **20**, 4137–4144.

Nehlin,J.O. and Ronne,H. (1990) *EMBO J.*, **9**, 2891–2898.

Nehlin,J.O., Carlberg,M. and Ronne,H. (1991) *EMBO J.*, **10**, 3373–3377.

Parkhurst,S.M., Harrison,D.A., Remington,M.P., Spana,C., Kelley,R.L., Coyne,R.S. and Corces,V.G. (1988) *Genes Dev.*, **2**, 1205–1215.

Pavletich,N.P. and Pabo,C.O. (1991) *Science*, **252**, 809–817.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sharma,K.K. and Arst,H.N.,Jr (1985) *Curr. Genet.*, **9**, 299–304.

Smith,D.B. and Johnson,K.S. (1988) *Gene*, **67**, 31–40.

Sophianopoulou,V. and Scazzocchio,C. (1989) *Mol. Microbiol.*, **3**, 705–714.

Sophianopoulou,V., Suárez,T., Diallinas,G. and Scazzocchio,C. (1993) *Mol. Gen. Genet.*, **236**, 209–213.

Yu,J., Donoviel,M.S. and Young,E.T. (1989) *Mol. Cell. Biol.*, **9**, 34–42.

Received on May 21, 1993; revised on October 15, 1993