ELSEVIER

FEBS Letters 342 (1994) 43–48

FEBS LETTERS

FEBS 13832

In vitro binding of the two-finger repressor CreA to several consensus and non-consensus sites at the *ipnA* upstream region is context dependent

Eduardo A. Espeso, Miguel A. Peñalva*

Centro de Investigaciones Biológicas del CSIC, Velázquez 144, Madrid, 28008, Spain

Received 21 January 1994; revised version received 11 February 1994; accepted 16 February 1994

Abstract

The two zinc-fingers of the *Aspergillus nidulans* repressor CreA recognize the consensus hexanucleotide 5'-SYGGRG-3'. We have determined all the CreA binding sites in a ~2 kb region upstream the *ipnA* gene. Our analysis shows that (i) CreA binds to certain consensus sites in a context-dependent manner; (ii) five non-consensus 6-bp sequences are also recognized by CreA; this non-canonical binding correlates with the presence of a second, neighbouring CreA binding site, suggesting that recognition of two linked sites stabilizes CreA binding. Our results suggest that the binding possibilities of CreA might be more complex than originally envisaged.

Key words: Transcription factor; DNA binding; Zinc finger; *Ascomycetes*; Carbon catabolite repression

1. Introduction

The CreA/MIG1 group of transcriptional repressors mediates carbon catabolite repression in the ascomycetes *Aspergillus nidulans* and *Saccharomyces cerevisiae* [1–4]. These transcriptional regulators recognize their cognate sequences in the promoters via DNA binding domains (DBDs) with two canonical C_2H_2 zinc-fingers. The almost complete identity between the amino acid sequences of the MIG1 and CreA zinc-finger regions [4] indicates that they recognize a rather similar DNA sequence motif.

Analysis of the co-crystal structure of the three finger Zif-268 DBD bound to its DNA target [5] has shown that each finger recognizes a contiguous base triplet, using R or H side chains to bind guanines in the DNA. These R or H residues are located in only two of three possible base contact positions in each finger (positions s3, s6 and m3; nomenclature of Jacobs [6]). Therefore, only two of the three bases in a given triplet of the Zif-268 target sequence are bound by the corresponding zinc-finger.

In contrast to the situation in Zif-268, the first of the two MIG1/CreA fingers contains H or R residues in the three possible base contact positions. The consensus sequence for CreA binding derived from DNase I protection analysis of the *alcR* and *alcA* promoter regions was 5'-SYGGGG-3' [7]. This sequence is also present in four sites protected from DNase I digestion by MIG1 [3]. These data suggested that the first CreA zinc finger could recognize three guanines (the second base triplet of the consensus) in the DNA. Cubero and Scazzocchio [8] have confirmed this prediction after analysis of the CreA binding sites at the *A. nidulans prnD-prnB* intergenic region. In addition, they showed that a 6-bp sequence 5'-GCGGAG-3' is a functional CreA binding site, validated by the previously characterised *cis*-acting mutation prn^d-20 [9]. Therefore, they expanded the consensus recognition sequence to 5'-SYGGRG-3° [8]. Binding to SYGGAG sequences is difficult to reconcile with the Zif-268 model.

We present in this paper the study of the in vitro CreA binding sites present at the *A. nidulans ipnA* gene upstream region. Although most likely this gene is not under *creA* control [10] the high G+C content found in certain subfragments of this region results in a high frequency of in vitro CreA binding sites. Our analysis shows that CreA recognizes, in addition to consensus SYGGRG sequences, other non-consensus 6-bp sites. Binding to the latter sites (and to certain consensus ones) is context dependent.

2. Materials and methods

2.1. Strains and plasmids

Escherichia coli DH5αF' [F'*lendA1, hsdR17, supE44, thi-1, recA1, gyrA*(Nal^r), *relA1, D(lacZYA-argF)U169 (f80dlacD(lacZ) M15*] was used for cloning experiments, plasmid maintenance and expression of the GST::CreA fusion.

Subfragments of the *ipnA* upstream region (nucleotide sequence in [11]) were obtained from plasmids whose inserts derived from λEMBL4 recombinant phages containing the *ipnA* gene, isolated by Ramón et al.

* Corresponding author. Fax: (34) (1) 5627518.

SSDI 0014-5793(94)00217-J

[12] from a genomic library of *A. nidulans biA1, veA1*. All recombinant plasmids were derivatives of pUC18 or pBS-SK+. DNA manipulations were made using standard procedures [13].

2.2. Obtention of GST::CreA(35–240) fusion protein*

Plasmid pGEX-*creA*, obtained from B. Felenbok, was constructed from pGEX-2T (Pharmacia). It encodes a fusion between gluthatione *S*-transferase (GST) and amino acids 35–240 of the CreA polypeptide, followed by an extra tail of seven amino acids codified by vector sequences [7]. The fusion protein was obtained in *E. coli* and purified by glutathione sepharose affinity chromatography essentially as described in [14], using 50 mM Tris-HCl, pH 8.0, 2 mM DTT and 5 mM glutathione (reduced form) as elution buffer. Glutathione was removed after filtration through a Sephadex G50 spun column (5 ml bed volume per 0.5 ml of eluted GST::CreA) equilibrated with storage buffer (50 mM Tris-HCl, pH 8.0, 5 mM DTT and 15% (v/v) glycerol). The columns were centrifuged at 2,000 rpm and 4°C for 2 min in a Sorvall HB4 rotor. The protein was stored at −80°C for months without detectable loss of activity.

2.3. Mobility shift assays

Restriction fragments of the *ipnA* upstream region were purified from polyacrylamide gels [15] and end-labelled after filling-in the protruding ends with Klenow fragment in the presence of [α-^{32}P]dCTP (3,000 Ci/mmol). The specific activity of the labelled fragments was usually 2.5×10^6 cpm/pmol. Mobility shift experiments were made as described [16]. Target fragments (1 ng, usually 20,000 cpm) were incubated in a buffer containing 25 mM HEPES-KOH, pH 7.9, 50 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 1mM DTT and 20% glycerol with 1 μg of purified GST::CreA(35–240*), in the presence of 3 mg of non-specific competitor (poly(dI-dC), Pharmacia). Reaction mixtures were kept in ice for 30 min and loaded onto a 5% polyacrylamide gel in 0.5 × TBE. Gels were run in the cold at 200 V, dried and exposed to X-ray film.

2.4. DNase I footprinting assays

Strand-specific labelling of fragments used in DNase I footprinting experiments was as follows: a plasmid containing an appropriate subfragment was cut open with a restriction enzyme and end-labelled as described above. The labelled plasmid was purified by spun-column chromatography using Sephadex G50 and restricted with a second enzyme which released the relevant fragment. The digestion was run on a polyacrylamide gel and the relevant labelled fragment was identified by autoradiography and eluted as in [15].

Binding reactions, carried out exactly as described above, were treated with 0.5 units of DNase I (Pharmacia, FPLC pure, diluted in 10 mM Tris-HCl, pH 7.5, 10 mM $CaCl_2$, 10 mM $MgCl_2$ and 10% glycerol) for 2 min at room temperature. Reactions were stopped after addition of 6 ml of a solution containing 5 M ammonium acetate and 2 mg/ml tRNA. Samples were phenol and chloroform extracted, precipitated with ethanol and the pellets were taken up in a small volume of DNA sequencing loading buffer. The ladders were resolved in a 6% polyacrylamide/7 M urea sequencing gel, alongside with a G+A Maxam and Gilbert cleavage [15] of the fragment as a standard, which was used to identify nucleotide positions.

3. Results and discussion

The *ipnA* upstream region contains several stretches of high G+C content. Due to this base composition, the occurrence of 6-bp sequences which match or are very close to the consensus hexanucleotide 5'-SYGGRG-3' recognized by the CreA DBD is relatively frequent. We have taken advantage of this library of putative CreA sites to investigate (i) if all sequences matching the consensus actually bind CreA, and (ii) if all sequences which bind CreA conform to the proposed consensus.

A 1,948 bp region upstream *ipnA* was conveniently subdivided in a set of 9 overlapping fragments (see Fig. 1) of very similar sizes. Each ^{32}P-labelled fragment was incubated with an excess of GST::CreA(35–240*) protein (to maximise the probability of detecting even weak CreA binding sites) and the formation of com-

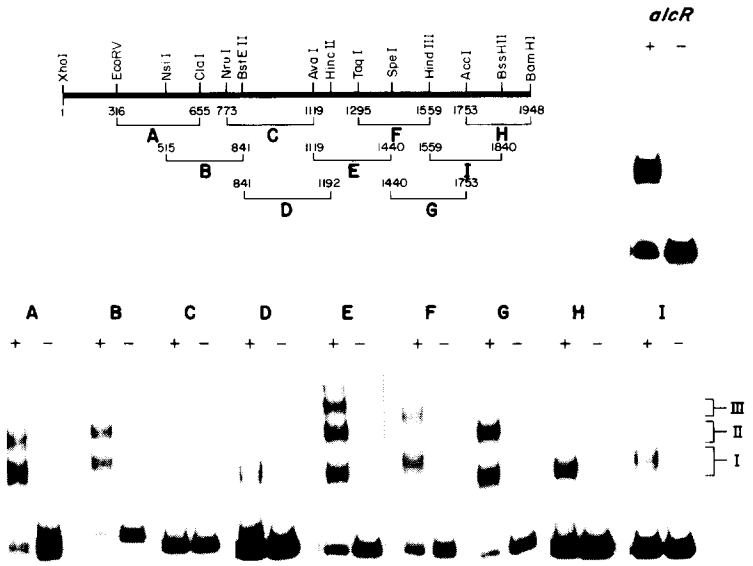

Fig. 1. Mobility shift assays using fragments of the *ipnA* promoter. A region of 1948 bp (from a *Xho*I site located at −2,004 bp from the initiation ATG to a *Bam*HI site at −56) was divided in a set of overlapping fragments of similar size. These fragments were end-labelled and individually tested in mobility shift assays in the presence of GST::CreA(35–240*) (+ symbols on top of the corresponding lanes) or in the presence of an equivalent amount of purified GST (− symbols). The limits of the fragments are shown using the *Xho*I site as origin of coordinates. No binding was detected upstream the A fragment. As a positive control, identical reactions (shown on the right) were carried out using a 340 bp *Tth*111I–*Ava*I fragment from the *alcR* promoter which contains the CreA binding site *alcR*.A 7. Also indicated on the right is the position of retardation complexes of similar size formed with the different fragments of the *ipnA* promoter, labelled from I to III in decreasing order of mobility. The retardation complex formed by fragment H corresponds to an artificial site created after fragment subcloning and has not been considered in this work.

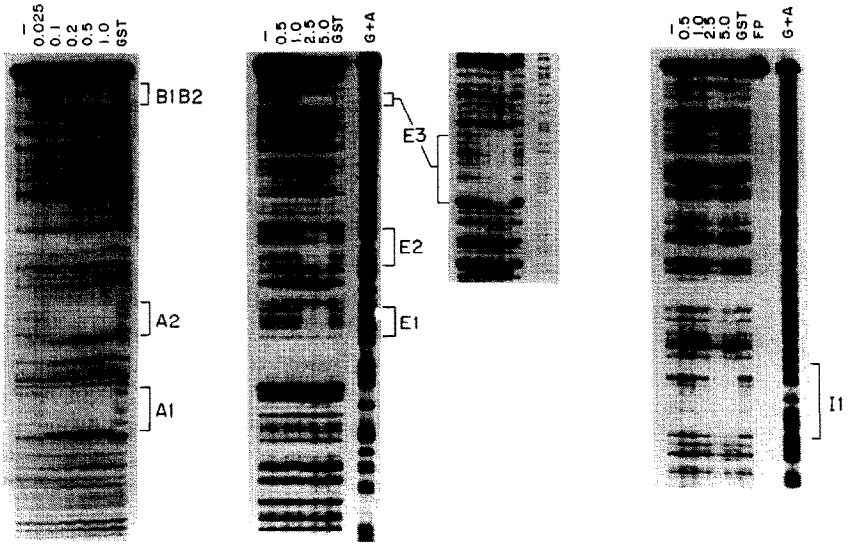

Fig. 2. DNAse I footprinting analysis of CreA binding sites at the *ipnA* promoter. Representative footprints for sites in fragment A (non-coding strand), E (non-coding strand for sites E1 and E2; coding strand for site E3) and I (non-coding strand) are shown. The footprint corrresponding to sites B1 and B2, which can be clearly distinguished on the upper part of the A ladders, has been published elsewhere 16 and is not included for simplicity. Fragments were incubated with different amounts of the fusion protein, as indicated on top of the corresponding lines. The experiments include controls carried out in the absence of protein (−) or in the presence of 5 μg of purified glutathione thiotransferase (GST). A Maxam and Gilbert G+A reaction was run in parallel to identify nucleotide positions. The protected regions corresponding to GST::CreA binding sites are indicated (see Fig. 3 for details).

plexes analised by gel electrophoresis (Fig. 1). Fragment E forms three complexes of different mobilities with GST::CreA(35–240*), fragments A, B, F, and G form two complexes each and fragments H and I form just one complex. Other fragments do not bind or bind CreA very weakly.

These results can be rationalized as follows: complex I (the highest mobility complex in all cases) represents the binding of one molecule of fusion protein to a DNA fragment, complex II results from simultaneous binding of two molecules, and complex III (fragment E) represents simultaneous binding of three molecules of protein to the fragment. Therefore, the number of retardation complexes formed by a fragment (see Fig. 1) would coincide with the actual number of CreA binding sites it contains (but see also below). DNAse I footprinting experiments (Fig. 2) showed that (with the exception of sites A1 and A2, see below) this interpretation was correct and served to confirm that protected regions contain sequences related to the CreA consensus binding site. The precise location of the eight binding sites present in fragments A, B, E and I and the nucleotide sequences of the regions protected from DNase I digestion are shown in Fig. 3 and will be discussed in the next paragraphs. An additional binding site is present at the 119-bp region located between fragments E and I. This site results in the formation of two retardation complexes with fragments F (which also contains E3) and G (which also contains I1). The only sequence in this region which is related to the CreA consensus site is 5'-taTTGGGG, which starts at coordinate 1525. This site has not been analysed by DNase I footprinting, as it is identical (including nucleotides at −1 and −2) to B1.

3.1. *Sites which conform to the consensus 5'-SYGGRG-3' and bind CreA*

These are sites E3 (the highest affinity site in fragment E, see Fig. 2) and I1 (the only site in this fragment). The 5'-GTGGGG-3' CreA binding site in E3 is identical to sites 4.2 and 2 in the *prn* intergenic region [8] and to site *alcR*.B in the *alcR* promoter [7]. These four sites bind CreA in vitro despite of their different sequence context. Site I1 (5'-GCGGGG-3') is identical to sites *alcR*.A and .C, to sites *alcA*.A and .B and to *prn*.4.1. Again, these 6 CreA binding sites have a different sequence context. The context independence of sites belonging to the GYGGGG subclass could be explained by assuming that CreA binds to these sites in a typical Zif-268 manner, with base contacts as previously proposed by others [8].

3.2. *Consensus hexanucleotides which do not bind CreA*

Fragment E contains the sequence GCGGAG (Table 1). CreA binding to this sequence is almost undetectable (Figs. 1 and 2), despite the fact that it perfectly matches the *prn*.3.2 site, used to propose the extended consensus sequence recognized by the CreA DBD [8]. This result shows that the binding ability of this sequence is context dependent. Both the *prn* and the *ipnA* sequences have an adenine at position −2, but remarkably, the *ipnA* sequence contains a guanine at position −1, whereas *prn*.3.2 has a thymine. Methylation of A(−2) and of the adenine complementary to the T (−1) interferes with

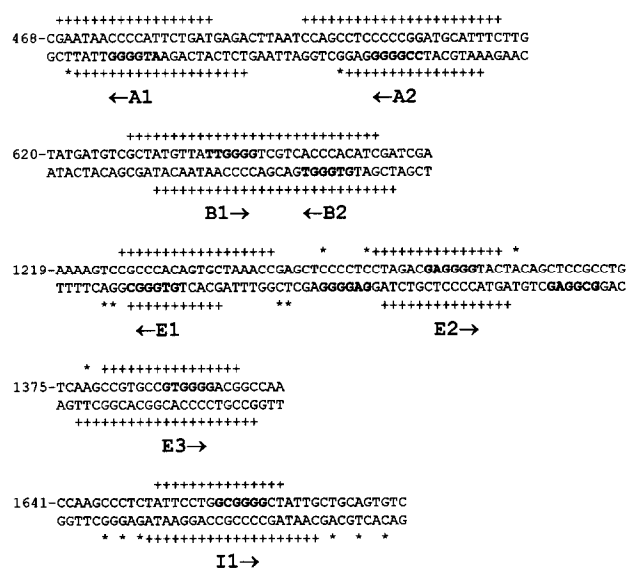

Fig. 3. Binding sites for GST::CreA(35–240*) at the *ipnA* upstream region. The nucleotide sequences (upper strand is the coding one for *ipnA*) which include the binding sites determined in this work (see Fig. 2) are shown in detail. Numbers on the left indicate nucleotide positions relative to the *Xho*I site located at −2,004 bp from the start of translation of *ipnA* (see Fig. 1). Positions in each strand which are protected from DNase I digestion are marked with + symbols. Asterisks indicate hypersensitivity to DNase I. The hexanucleotides (G-rich strand) which represent the core of the CreA binding sites are shown in bold.

CreA binding to *prn*.3.2 [8]. Therefore, the presence of a G at −1 might explain the lack of binding ability of the *ipnA* sequence.

The binding ability of the sequence CCGGGG has not been previously investigated. Two of these sequences are present in the *ipnA* upstream region (Table 1) and show absolute context dependence for CreA binding. While the CCGGGG sequence in fragment C does not bind CreA (Fig. 1), the second sequence corresponds to the A2 binding site (Figs. 1–3) which is the highest affinity binding site of all sequences investigated in this work. According to the Zif-268 model [5], the second CreA finger would only make a canonical contact with the G of the first base triplet of the A2 site. However, the context dependence of the binding to the A2 hexanucleotide strongly suggests that other base contacts (perhaps to the adenines at base pairs −1 and −2) may stabilize the interaction of this second finger. In that case, the sequence recognized by the second finger would extend beyond one base triplet.

In agreement with the importance of certain nucleotides at positions −1 and −2, all MIG1 binding sites which have been described have the A(−2) T(−1) sequence [3]. Base pairs −1 and −2 (relative to the three base triplets) were not included in the co-crystals of the Zif-268 complex [5], indicating that these contacts are not essential for high-affinity binding to DNA by three-fingered proteins. This might not be the case with two-fingered proteins, such as CreA. Indeed, the rules deduced from the Zif-268 complex are not necessarily universal. For example, the structure of the two-fingered Tramtrack DBD-cognate DNA complex has revealed novel features of the zinc finger–DNA interaction [17].

In summary, while binding of the CreA DBD to GYGGGG sequences (sites E3 and I1) is context independent, binding to the 'four guanines' site CCGGGG shows absolute context dependence and apparently correlates with the presence of T at −1 and A at −2 of the two base triplets recognized by the zinc fingers. A similar conclusion can be drawn by contrasting the binding properties of the only GCGGAG sequence investigated here to those of site *prn*.3.2 [8].

3.3. *Non-consensus hexanucleotides which bind CreA*

We have detected five CreA binding sites (A1, B1, B2, E1 and E2, see Fig. 3) which do not fit within the SYGGRG consensus. Site A1 corresponds to the non-consensus sequence ATGGGG. It is located at 30 bp (roughly three integral helical turns of B-DNA) from the high-affinity binding site A2 (see Fig. 3). The latter satisfies the theoretical conditions for high-affinity CreA binding: it conforms to the SYGGRG consensus [8] and has a thymine at −1 and an adenine at −2 (see above). Indeed, site A2 has a much higher affinity for GST::CreA than any other site in fragments B, E and I, as shown by DNase I footprinting experiments with different amounts of protein (see Fig. 2). Strikingly, the affinity of the non-consensus site A1 is not distinct from that of A2. This would be as expected if binding to sites A1 and A2 were ligated phenomena (see below).

Fragment A forms a high proportion of complex I, representing binding of one protein molecule (Fig. 1). This occurs despite of the fact that fragment A contains,

Table 1
Sequences which show context dependence for CreA binding

Binding site * sequence	Binding ability**	Site desig.
Sequences belonging to the SYGGRG consensus		
1,028'-gcgCCGGGG	−	
513' -catCCGGGG	+++	A2
tatCCGGGG	MIG1 binding [2]	*SUC2* B
1,283'-cagGCGGAG	−	
tatGCGGAG	CreA binding [8]	*prn* 3.2
Sequences outside the SYGGRG consensus		
1,163 -cctGAGGGG	−	
1,257'-tagGAGGGG	−	
1,258 -gacGAGGGG	+	E2

* Sites belonging to the *ipnA* upstream region are numbered from *Xho*I at −2,006 from the ATG. A dash indicates non-coding strand (for *ipnA*).
** In the case of *ipnA* sites, the binding ability (this work) is indicated relative to other sites in this region; '−' indicates absence of or weak CreA binding; '+++' indicates protection to DNAse I digestion; '+' means less intense protection to DNAse I.

in addition to A1 and A2, at least the B1 site (and possibly a functional B2 site, as it includes the B2 hexanucleotide and the four bp from positions −1 to −4, see Fig. 3), suggesting that a single molecule of protein would mediate protection of sites A1 and A2. Consistent with this, fragment A forms this high proportion of complex I under conditions in which sites A1 and A2 are fully protected from DNase I digestion (see Figs. 1 and 2). These observations could be rationalised if the protein would behave as a dimer, with each of the two CreA DBDs simultaneously contacting sites A1 and A2. The conformational changes caused in the DNA by this mode of binding to sites A1 and A2 could explain the remarkable hypersensitivity to DNase I detected in positions 5 and 4 after the respective 6-bp sites (see Figs. 2 and 3).

Fragment B contains B1 and B2, a pair of closely linked, non-canonical binding sites (Fig. 3). These divergent (head to head orientation) sites are separated by 5 bp. The extension of the DNase I footprint in this region strongly suggests that the DNA sequence containing B1 and B2 is protected by two CreA DBDs (Fig. 3, compare for example to that in sites E3 or I1). The possible dimeric nature in solution of GST::CreA(35–240*) mentioned above might explain this footprint. However, mobility shift experiments with fragment B (Fig. 1) showed the formation of two retardation complexes of very different mobility. This indicates that each site of the B1B2 pair binds a different molecule of GST::CreA(35–240*), with the low mobility complex corresponding to simultaneous binding of two molecules of protein to each of the B sites and the high mobility complex representing single occupancy of just one site (B1 or B2). Finally, the high proportion of low mobility complex formed with fragment B (Fig. 1) suggests that close linkage of the non-consensus sites B1 and B2 stabilizes CreA binding.

Divergent sites E1 and E2 are arranged in a tail-to-tail orientation and separated by 28 bp. These sites in fragment E are weaker binding sites than E3 (Figs. 1 and 2). Site E2 contains the sequence GAGGGG (Table 1). The same sequence occurs 5 bp upstream at the other strand (Fig. 3) but is not protected from DNAse I digestion, indicating that binding to GAGGGG is context dependent (see Table 1). Fragment E, with three binding sites revealed by DNase I footprinting experiments, forms three retardation complexes (Fig. 1). This strongly suggests that each site binds a protein molecule. Therefore, as discussed above for B sites, the occurrence of the neighbouring, otherwise weak binding sites E1 and E2, apparently favours binding of GST::CreA. Binding to sites E1 and E2 results in hypersensitivity to DNase I digestion in the linker DNA (Fig. 3), perhaps as a result of the interaction between bound protein molecules.

In summary, the GST::CreA(35–240*) protein used here binds to five hexanucleotides, represented by sites A1, B1, B2, E1 and E2, which do not fit within the previously proposed consensus SYGGRG [7–8]. Binding to all of these atypical sites coincides with the presence of a second, closely linked CreA binding site in their proximities. This absolute correlation suggests that binding of CreA to closely located sites stabilizes the resulting complexes. This conclusion could be simply explained if the GST::CreA(35–240*) fusion used here would be a dimer in solution, perhaps through the GST moieties [18]. Our results (see above) strongly suggest that a fraction of the high mobility complex formed with fragment A results from binding of a dimeric protein which simultaneously contacts A1 and A2. Changes in DNA conformation imposed by this way of binding would explain the remarkable hypersensitivity to DNase I detected in certain positions near the sites.

However, the apparent dimeric nature of GST::CreA-(35–240*) in solution cannot explain by itself the pattern of interaction with sites B1 and B2 or E1 and E2, all of which do not fit within the SYGGRG consensus. In these cases, the complexes formed are consistent with binding of one protein molecule to each site of these pairs. This supports the possibility that the presence of adjacent sites could stabilize CreA binding, thereby providing a mechanism to increase the binding specificity of the two-fingered protein CreA. In this context, it is noteworthy that the only three in vitro CreA binding sites which have been shown by others to be functional in vivo are closely linked to other binding sites: *alcR*.A is located 15 bp upstream of the high-affinity binding sequence GCGGAG [7], *alcA*.B is located 25 bp upstream of a divergent consensus CTGGGG sequence [19], and the two 6-bp divergent sites to which the carbon catabolite derepressed mutations *prn*^{d}-20 and *prn*^{d}-22 map are separated by 1 bp [8].

Acknowledgements: We thank B. Cubero and C. Scazzocchio for communicating their results prior to publication to B. Felenbok for plasmid pGEX-CreA and a plasmid clone containing site *alcR*.A. This work was supported by the Commission of the European Communities through BRIDGE contract BIOT-CT90–0169 and by the Comisión Interministerial de Ciencia y Tecnología (Spain) through Grant BIO 91–0671. E.E. is holder of a predoctoral fellowship from the Programa Nacional de Formación del Personal Investigador (M.E.C.). We also thank E. Reoyo for technical assistance, and R. Uña and A. Hurtado for photographic and delineation artwork, respectively.

References

[1] Bailey, C.R. and Arst Jr., H.N., (1975) Eur. J. Biochem. 51, 73–577.
[2] Nehlin, J.O. and Ronne, H. (1990) EMBO J. 9, 2891–2898.
[3] Nehlin, J.O., Carlberg, M. and Ronne, H. (1991) EMBO J. 10, 3373–3377.
[4] Dowzer, C.E.A. and Kelly, J.M. (1991) Mol. Cell. Biol. 11, 5701–709.
[5] Pavletich, N.P. and Pabo, C.O. (1991) Science 252, 809–817.
[6] Jacobs, G.H. (1992) EMBO J. 11, 4507–4517.
[7] Kulmburg, P., Mathieu, M., Dowzer, C., Kelly, J. and Felenbok, B. (1993) Mol. Microbiol. 7, 847–857.

[8] Cubero B. and Scazzocchio, C. (1994) EMBO J. 13, 407–415.
[9] Sophianopoulou, V., Suárez, T., Diallinas, G. and Scazzocchio, C. (1993) Mol. Gen. Genet. 236, 209–213.
[10] Espeso, E.A., Tilburn, J., Arst Jr., H.N. and Peñalva, M.A. (1993) EMBO J. 12, 3947 3956.
[11] MacCabe, A., van Liempt, H., Palissa, H., Unkless, S., Riach, M., Pfeifer, E., von Döhren, H. and Kinghorn, J. (1991) J. Biol. Chem. 256, 12646–12654.
[12] Ramón, D., Carramolino, L., Patiño, C., Sánchez, F. and Peñalva, M.A. (1987) Gene 57, 171–181.
[13] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning. A laboratory manual, Cold Spring Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[14] Kulmburg, P., Sequeval, D., Lenouvel, F., Mathieu, M. and Felenbok, B. (1992) Mol. Cell. Biol. 12, 1932–1939.
[15] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[16] Pérez-Esteban, B., Orejas, M., Gómez-Pardo, E. and Peñalva, M.A. (1993) Mol. Microbiol. 9, 881–895.
[17] Fairall, L., Schwabe, J.W.R., Chapman, L., Fintch, J.T. and Rhodes, D. (1993) Nature 336, 483–487.
[18] Mannervik, B. (1985) Adv. Enzymol. 57, 357–417.
[19] Hinz, W.E. and Lagosky, P.A. (1993) Biotechnology 11, 815–818.