CHARACTERISATION OF DNA-BINDING BY THE CREA PROTEIN OF *ASPERGILLUS NIDULANS*

by

Nicole D. Chamalaun-Hussey B.Sc. (Hons)

Thesis submitted for the degree of
Doctor of Philosophy
to
The University of Adelaide
(Department of Genetics)

June 1996

This copy is printed on archival paper

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, to be available for loan and photocopying.

Signed:.. [redacted] . Date:..26/6/96..

Acknowledgements

I am extremely grateful to my supervisor Dr. Joan Kelly for her patience, skill, knowledge and for all the endless help she has given me over the years.

I would like to thank Prof. Michael Hynes, Dr. Meryl Davis and all the gang at Melbourne University, Genetics Department (especially Richard Todd and Rachael Murphy) for making available methods, *facB* and *amdS* clones, oligonucleotides and for sharing the sequence of *facA* and *facB* before publication. A special thankyou to Prof. Michael Hynes for communicating unpublished results.

I am indebted to many people, past and present from the Genetics Department for social diversions and for their technical help. Robin Lockington, Robert Shroff, Rosemary Sladic, Sue O'Connor, Michael Ayliffe, Sharon Orford, Jamie Brady, Mary Drysdale, Celia Dowzer, Angela Binns and all the honours students for making my life look good. I would like to thank Doug Pottrell, Roger Brock and Paul Moir for their help with all the gadgets that seem to endlessly breakdown around the department. Expert technical assistance from the CSU staff, Sharon Kolze and Michelle Adamou is gratefully acknowledged. I would also like to thank other members of the Genetics Department staff for their help over the years, Jeremy Timmis, Ruth Rofe and Carolyn Leach (for letting me tutor), Rory Hope (for taking such good care of Damian), Marelle Smith, Helena Richardson and Rob Saint. I'd especially like to thank Robin Lockington and Rob Shroff for proof reading and general coming to the rescue when things went wrong.

I'd like to thank my family and friends for putting up with me over the last few years (It hasn't always been easy, has it!). Especially I'd like to thank my father Dr. Francois Chamalaun for the use of his computing facilities, his invaluable advice (stop procrastinating), and help and my mother Tilly Chamalaun for all the long hours and terrible pay that goes with being a mum. Thanks guys for all your love and support, this thesis would not have materialised if it weren't for you. A special thanks to Toshika Chamalaun for help at the eleventh hour.

Last but not least, I would like to thank my husband, Damian Hussey for his love, help and support and all the "Husseys" for being such great in-laws.

Abstract

In *Aspergillus nidulans* carbon catabolite repression is mediated by the negatively acting wide domain regulatory gene *creA* and control has been shown to be at the level of transcription for many genes. The theoretically translated protein sequence contains a Cys_2His_2 zinc finger motif with striking similarity to the zinc fingers of the MIG1 protein involved in carbon catabolite repression in *Saccharomyces cerevisiae* (Dowzer and Kelly, 1991; Nehlin *et al.*, 1990).

In this study a GST-CREA fusion protein was shown to bind to DNA fragments from the 5' region of three genes where there existed evidence for *creA* control. Two of these were the structural gene for acetamidase, *amdS*, and the positive regulatory gene, *facB*, involved in acetate induction of *amdS*. In addition, binding of CREA to the 5' region of *creA* was found and confirms an autoregulatory role for the *creA* gene product. The binding sites within the 5' regions of *amdS* and *facB* were further localised by DNase I sensitivity assays. Protected regions spanned 5' G/C T/C G G G G 3' sequences and sequences which differed only slighly from this. This was very similar to the binding site for MIG1 which has a consensus of 5' A/T A/T A/T A/T T G/C T/C G G G G 3'.

In order to determine more precisely preferences for the core binding site and 5' extended context, a CASTing (cyclic amplification and selection of targets) experiment was undertaken. This technique would also have allowed the determination of a totally novel binding site for the CREA protein had one existed. Results from this analysis showed a preference for a T in position 5 (where the most 3' G is position 1) and a G in position 6. A T in position 7 was present in 75% of all selected sequences. In contrast to the MIG1 consensus an AT rich region immediately 5' of the core recognition sequence was not found, although this region was generally AT rich. A C in position 22 occured at high frequency and may form part of a recognition element with sequences from within the primers. Evidence was found to suggest that the variation 5' G T/C G G G T 3' was able to bind CREA when located in close proximity to other sites. The presence of additional binding sites was found to increase binding affinity.

To enable detection of native CREA protein two sets of polyclonal antibodies were raised. One set was raised against the amino terminal portion of *in vitro* generated protein and the other against the carboxy terminal region. When these sera were used in Western analyses of *E. coli* expressed proteins they were able to detect CREA. These antibodies were, however, insufficiently specific to allow detection of native CREA. Mobility shift analysis using nuclear extracts were undertaken. Binding to the fragment containing the CREA binding sites 5' of *amdS* by wild type extracts and extracts from a mutant strain which produces a truncated CREA protein that contains the zinc finger region was found. Proof that this binding was due to CREA awaits the generation of antibodies with higher specificity.

Table of Contents

Errata

Line 25 on page 88 should read:

show increased sensitivity to DNase I digestion (ie are hyperdigested). Hyperdigested bases result from a change in conformation of the DNA molecule by a bound CREA molecule(s), making some bases more susceptible to digestion by DNase I. Although a very.....

Line 24 on page 128 should read:

disrupted within this fragment was the **A4** site. Although site **A4** is disrupted, two new potential sites for CREA binding have been formed (underlined below) by the insertion of linker sequence (shown in bold).

CTCCCCCCGG**GATC**CCGGGGATCA

Therefore binding by this fragment may reflect binding to these new sites instead of site **A4**. A fragment containing the.....

Additional Figure 1 (see text page 143):

5' *amdS*	TC GCCC G CTG GAG AGCATCCTG AA TCGC A GTG GGG AAGCCACAC AT TGAT C CCG GGG GAGCGGATG GA TCAA T GAG GAG AATGAGGGG GG AGAA T GAG GGG GATGCGGGG GG GGGA T GCG GGG CTAAAGAAG
Consensus	**tgaa T GTG GGG aa** **c CA A** **C**

5' *facB*	CT AGAT T CTG GAG T CG AACA G CCG GGG AAGACCAGT GC CGAT T CCG GGG GAAAGGAAA
Consensus	**T CCG GGG** **A**

Between paragraphs 1 and 2 on page 279 should be the following:

Ouellette, M.M, Chen, J., Wright, W.E. and Shay, J.W. (1992) Complexes containing the retinoblastoma gene product recognize different DNA motifs related to the E2F binding site, *Oncogene*, **7**, 1075-1081.

List of Tables

List of Figures

Abbreviations

A:	Adenine
Ac:	acetate
ACV:	δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine
ATP:	adenosine-5'-triphosphate
bp:	base pair(s)
BSA:	bovine serum albumin
C:	cytidine
CAP:	catabolite gene activator protein
cAMP:	adenosine 3',5'-monophosphate (cylic AMP)
CASTing:	cyclic amplification and selection of targets
cDNA:	DNA complementary to mRNA
Ci:	Curie
c.p.s.:	counts per second
CRP:	*E. coli* cyclic AMP receptor protein
dATP:	2'-deoxyadenosine-5'-triphosphate
DMS:	dimethyl sulfate
DNA:	deoxyribonucleic acid
EDTA:	disodium salt of ethylenediaminetetraacetic acid
EGTA:	ethylene glycol-*bis*(β-aminoethyl ether)N,N,N',N'-tetraacetic acid
DTT:	1,4-dithiothreitol
g:	gram force of gravity
G:	guanine
GABA:	γ-amino-n-butyric acid
GST:	glutathione-S-transferase
HEPES:	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
HTH:	α-helix-turn-α-helix
IPTG:	isopropyl-β-D-thiogalactopyranoside
NADP:	nicotinamide adenine dinucleotide phosphate
^{32}P-dATP:	2'-deoxyadenosine-5'-triphosphate
PAG:	polyacrylamide gel
PAGE:	polyacrylamide gel electrophoresis
PCR:	polymerase chain reaction
PMSF:	phenylmethyl sulfonylfluoride
poly{dI-dC}:	poly-deoxy-inosinic-deoxy-cytidylic acid
PTS:	phosphotransferase system
R:	adenine (A) or guanine (G)
5' RACE	rapid amplification of cDNA ends

S:	guanine (G) or cytidine (C)
SDS:	sodium dodecyl sulfate
T:	thymine
TCA:	tricarboxylic acid cycle
TEMED:	N,N,N',N'-tetramethylethylenediamine
Tris:	Trizma Base (tris[hydroxymethyl]aminomethane)
W:	adenine (A) or thymine (T)
X-gal:	5'-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
Y:	thymine (T) or cytidine (C)

SI (Système International) units are used thoughout.

CHAPTER 1

Chapter 1

Introduction

Many genes in *Aspergillus nidulans* are under the broad regulatory control of single regulator genes termed "wide domain regulatory genes". The topic of this thesis, *creA*, is one such wide domain regulatory gene which is involved in carbon catabolite repression. Carbon catabolite repression, also referred to as glucose repression, results in the reduced synthesis of a range of enzymes and permeases involved in the utilisation of alternative carbon sources when a more readily metabolised carbon source, such as glucose, is present. This phenomenon allows the organism to utilise resources efficiently.

The ascomycete fungus *A. nidulans* is an ideal organism in which to study the phenomenon of carbon catabolite repression as it is able to utilise many different compounds as sole carbon sources. It is able to grow on defined media and the phenotype of many mutations can be readily scored due to it being haploid. *A. nidulans* has both a sexual and an asexual cycle and the absence of mating types theoretically allows inter-crossing of any strains. It has eight linkage groups which correspond to eight chromosomes that vary in size from 2.9-5.0 Mbp (Brody and Carbon, 1989).

Many genes have been shown to affect carbon catabolite repression in the yeast *Saccharomyces cerevisiae*. Some of these play a direct role in the regulatory cascade while others play a general role in the transcriptional process, either by interacting with the transcriptional machinery or by affecting chromatin structure. In *A. nidulans* one gene, *creA,* has been implicated by mutational analysis in the control of many others. This gene has been cloned and sequenced (Dowzer and Kelly, 1989; Dowzer and Kelly, 1991). A null mutation that lacks the *creA* gene as well as 5' and 3' flanking sequences results in severe morphological effects in haploid cells (Dowzer and Kelly, 1991). This thesis focuses on the mechanism of action of the *creA* gene product of *A. nidulans,* and its ability to control the regulated expression of three genes for which there exists evidence for control by CREA at the transcriptional level. This chapter is a

review of current knowledge on the control of gene expression by DNA binding proteins, carbon catabolite repression, and the regulation of some genes under *creA* control.

1.1 Regulation of gene expression

Control at the level of transcription initiation plays an important role in the regulation of cellular functions. There are two processes that are required for correct gene expression. Firstly, the formation of a transcription initiation complex containing RNA polymerase II to direct and initiate transcription. Secondly, the interaction(s) of single proteins or large multiprotein complexes that bind DNA and regulate the initiation of transcription. This first process utilises general transcription factors which form the transcriptional machinery required for transcription from all RNA polymerase II promoters (reviewed by Zawel and Reinberg, 1992). The second process involves special transcription factors, more correctly called regulatory proteins or *trans* acting factors, which bind to sites within the 5' region of genes thus allowing direct or indirect interactions with the transcriptional machinery to increase or decrease the frequency of initiation of transcription (Mitchell and Tjian, 1989).

The structure of regulatory proteins in eukaryotes is modular. For example certain regions of a protein might be involved in DNA binding while others are involved in protein-protein interactions with other proteins in a multimeric group. In addition to these, regulatory proteins may contain activation domains (reviewed by Ptashne and Gann, 1990; Triezenberg, 1995) or repression domains (Renkawitz, 1990), signals for nuclear localisation (Boulikas, 1993), and cofactor binding, inducer binding or dimerisation domains. It is often a complex interaction of *trans* acting factors with *cis* elements or with one another that controls the specific spatial and temporal expression of genes. In order to understand how genes are regulated it is important to know how regulatory proteins are able to recognise and bind specific DNA sequences.

1.1.1 The helix turn helix class of DNA binding motifs

Many DNA binding motifs have been discovered which enable proteins to bind DNA in a sequence specific manner. The first such motif characterised was the α-helix-turn-α-helix

(HTH), which was found within the two λ bacteriophage proteins CRO and λ repressor (CI) and the *E. coli* cyclic AMP receptor protein (CRP), sometimes called the catabolite gene activator protein (CAP). These proteins contain two helices separated by a sharp beta turn (reviewed by Pabo and Sauer, 1984). They bind DNA as dimers and their recognition sequences have approximate dyad symmetry. Crystallographic studies of the λ repressor (Jordan and Pabo, 1988) have revealed that the carboxy terminal helix (now called the recognition helix or helix 3) lies in the major groove of DNA and makes base specific contacts. The second helix (amino terminal helix) lies across the recognition helix and via hydrophobic interactions gives the recognition helix stability.

1.1.1.1 The HTH motif of *Drosophila* homeodomain proteins

The HTH of eukaryotes are variations of the prokaryotic theme. Sequence similarities of *Drosophila* homeotic proteins to the HTH of *E. coli* proteins led to the discovery of the homeodomain. This is a region of 61 amino acids with typically >90% identity at the amino acid level. This high level of conservation between homeodomains led to the cloning of genes regulating transcription in many organisms including mouse (McGinnis *et al.*, 1984) and human (Levine *et al.*, 1984). Homeodomains have also been found in yeast (Shepherd *et al.*, 1984) and the fungus *Ustilago maydis* (Schulz *et al.*, 1990). Most bind very similar recognition sequences surrounding a core TAAT sequence (reviewed by Scott *et al.*, 1989; Treisman *et al.*, 1992) and thus specificity of function is provided by subtle differences in binding affinities *in vivo* and by protein-protein interactions (reviewed by Hayashi and Scott, 1990).

Crystallographic analysis of the *Drosophila* Engrailed homeodomain bound to a DNA oligonucleotide containing the recognition sequence TAAT has confirmed the importance of the HTH motif in binding (Kissinger *et al.*, 1990). Helix 1 and 2 pack against each other in an antiparallel manner. They span the major groove but are too far away from the DNA to make any base contacts. They are important in stabilising protein folding and in the positioning of helix 3. The recognition helix (helix 3) lies in the major groove, and makes extensive contacts with the DNA bases and with the sugar-phosphate backbone. The hydrophobic face of helix 3

packs against helix 1 and 2 to form the interior of the protein and the hydrophilic face of helix 3 fits into the major groove. Amino acids outside of the homeodomain are also responsible for sequence specific contact with the DNA since the amino terminal arm of the homeodomain lies in the minor groove and contacts bases that are adjacent to those contacted by helix 3.

Some homeodomain proteins are able to bind two unrelated recognition sequences (Hoey *et al.*, 1988; Dearolf *et al.*, 1989). For the Paired protein, if a TAAT sequence is not present in the recognition sequence then the alignment of the recognition helix is closer to the prokaryotic HTH changing the specificity of the interaction quite significantly (Treisman *et al.*, 1989). At the yeast *MATα2* locus, mutants defective in binding one site but not the other were obtained (Porter and Smith, 1986).

1.1.1.2 Other HTH variations

As more DNA binding proteins are discovered (reviewed by Nelson, 1995) additional variations on the HTH domain of prokaryotes are being found, and two examples are the tethered and winged HTH motifs. An example of a tethered HTH is the Oct-1 POU1 domain. This consists of two domains, a classical HTH similar to the λ repressor and a POU homeodomain similar to homeodomain proteins. The tethering implies both recognition sites must be close together and may increase the thermodynamic properties for co-operativity between the two domains (Klemm *et al.*, 1994). One example of the winged HTH motif is the hepatocyte nuclear factor (HNF)-3/ fork head DNA binding domain where two loops stick out from either side of the HTH motif and make contacts primarily along the phosphate backbone (Clark *et al.*, 1993). Some proteins, for example LexA, contain just one wing (Fogh *et al.*, 1994).

1.1.2 The zinc containing class of DNA binding motifs

1.1.2.1 The classical zinc finger DNA binding motif

The zinc finger class of DNA binding motifs was discovered in the structure of the polymerase III transcription factor TFIIIA from the frog *Xenopus laevis* (Miller *et al.*, 1985). This Cys_2His_2 class of Zinc fingers is characterised by the protein sequence X_3-Cys-X_{2-4}-Cys-X_{12}-His-X_{3-4}-His-

X_4, where X is any amino acid. TFIIIA combines with TFIIIB and TFIIIC to assemble into a stable complex on the ribosomal 5S gene which is a target for RNA polymerase III transcription (Segall *et al.*, 1980; Shastry *et al.*, 1982). TFIIIA contains nine consecutive zinc finger motifs with each pair of cysteines (in the α helix) and histidines (in the β sheet) tetrahedrally coordinating a Zn^{2+} ion. The intervening amino acids loop out forming the "finger" part of the motif (Fig. 1.1a). Using ion chelators such as 1,10-phenanthroline, DTT and EDTA it has been shown that the presence of a Zn^{2+} ion is essential for binding activity and the addition of other divalent cations does not restore binding (Miller *et al.*, 1985). Cys_2His_2 zinc fingers have now been found in more than 200 regulatory proteins including ADR1 from yeast (Blumberg *et al.*, 1987), the developmental regulator encoded by *brlA* from *A. nidulans* (Adams *et al.*, 1988), the *Drosophila* Kruppel protein (Rosenberg *et al.*, 1986), the mammalian transcription factor Sp1 (Kadonaga *et al.*, 1986), the mouse immediate early protein Zif268 (Christy *et al.*, 1988), and the WT1 protein which is encoded by a gene that is mutated in Wilms' tumour cells (Call *et al.*, 1990).

X-ray crystallographic studies were carried out on Zif268 (also known as Krox-24, NGFI-A and Egr1) to determine the precise points of contact between the amino acids in the finger and the base pairs in DNA (Pavletich and Pabo, 1991). Zif268 contains three zinc fingers and an overall structure in agreement with NMR studies on other zinc finger proteins (Parraga *et al.*, 1988; Lee *et al.*, 1989). The three zinc fingers form a semicircular structure which fits into the major groove of the DNA helix. The DNA-protein complex exhibits periodicity with each finger contacting the DNA helix in the same overall way and this reflects the periodicity of the three base pair recognition site for each finger. Since Zif268 contains three zinc fingers this results in a total of 9 bp in the recognition sequence (5' GCG TGG GCG 3'). The first finger makes primary contacts with bases within the 3' triplet of the recognition sequence the second the middle triplet and the last finger contacts the 5' triplet resulting in an "antiparallel" arrangement. From an inspection of the amino acids which make critical contacts with the DNA it was clear that arginine-guanine contacts play a critical role (Fig. 1.1b; Pavletich and Pabo, 1991). The Zif268 fingers recognise extremely GC rich sequences and this is also the case for the SP1 (Kadonaga

Figure 1.1a Folding scheme for the Cys_2His_2 zinc finger DNA binding motif

Folding scheme for the zinc finger motifs of TFIIIA from *Xenopus laevis* (taken from Miller *et al.*, 1985). Each Zn^{2+} metal ion co-ordinates two cysteines and two histidines. The amino terminal to carboxy terminal direction is shown with arrows. The circled residues are highly conserved and the positions shown with black circles were hypothesised by Miller *et al.*, (1985) to most likely contain residues with DNA binding side chains.

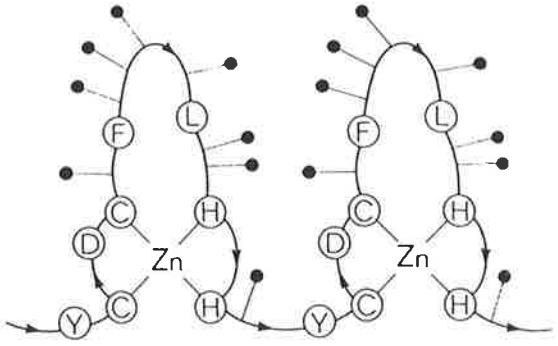
Figure 1.1b Crystal structure of the Cys_2His_2 zinc finger DNA binding motif from Zif268

1) Stereo diagram of the crystal structure of a peptide containing the zinc finger region of Zif268 bound to an oligonucleotide of its recognition site. Shown are the backbone atoms for residues 3- 87 and the side chains for some amino acids including those that contact the DNA base pairs (R18, D20, R24 of finger I; R46, D48, H49 of finger II; R74, D76, R80 of finger III). Zinc ions are shown as circles (taken from Pavletich and Pabo, 1991).

2) Sketch summarising the information in 1) above.

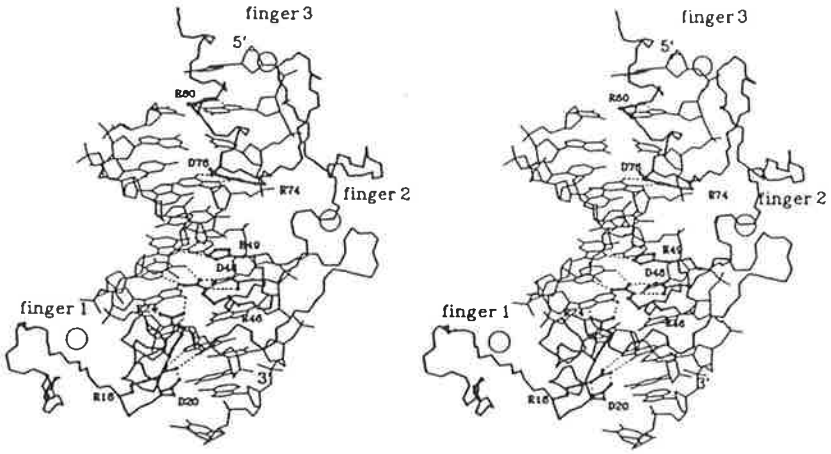
3) Sketch of the critical contacts made by amino acid side chains of the Zif268 peptide with its recognition site (from Pavletich and Pabo, 1991). Finger I is the amino terminal finger of the peptide and contacts the 3' end of the recognition site. These interactions are discussed further in section 3.8.

a)

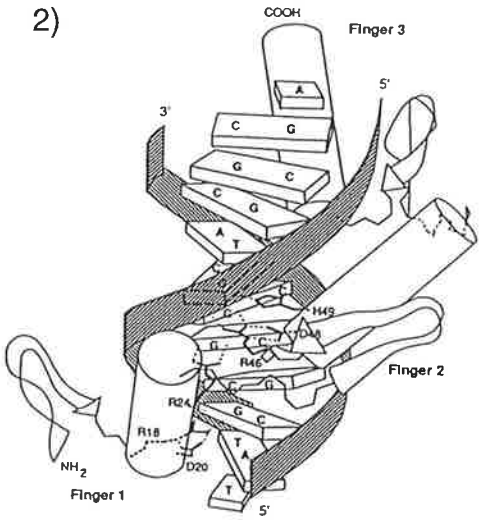

F
L
C
H
D
Zn
Y
C
H

b)

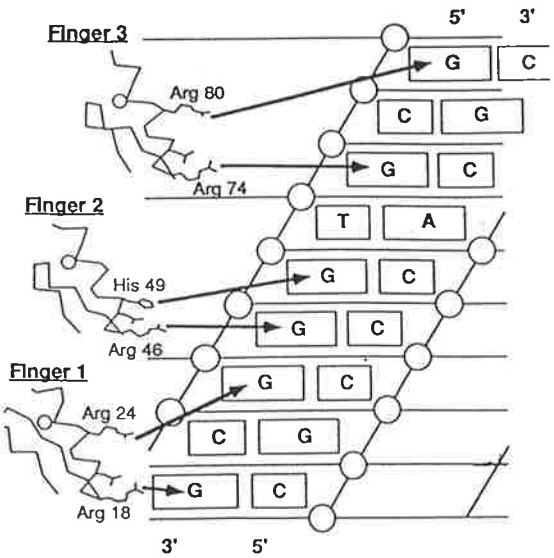
1)


finger 3
5'
finger 2
finger 1
3'

2)


COOH
Finger 3
5'
3'
Finger 2
Finger 1
NH2
5'

3)


5'
3'
Finger 3
Arg 80
Arg 74
Finger 2
His 49
Arg 46
Finger 1
Arg 24
Arg 18
3'
5'

et al., 1986) which has similar amino acids in the critical positions. Thus it appears that it may be possible to predict a proteins' recognition site from its amino acid sequence and this will be discussed in more detail in section 3.8. Further stabilisation of the protein-DNA complex results from interactions between the α-helix and β-sheet of Zif268 and the phosphate backbone of the DNA molecule.

1.1.2.2 The Cys_4 zinc finger motif

A variation on the classical zinc finger motif is the Cys_4 (Cys-X_2-Cys-X_{17}-Cys-X_2-Cys) zinc finger. Some proteins, including the vertebrate GATA family (so called because they bind WGATAR sequences, reviewed by Orkin, 1992) and ELT-1 from *Caenorhabditis elegans*, contain two Cys_4 fingers (Spieth *et al.*, 1991) whereas the fungal NIT2, GLN3 and AREA proteins contain a single Cys_4 finger (Fu and Marzluf, 1990a; Minehart and Magasanik, 1991; Kudla *et al.*, 1990). In the two fingered proteins the carboxy terminal finger is required for binding whereas the amino terminal finger provides stability and specificity of binding (Martin and Orkin, 1990; Yang and Evans, 1992). In single fingered proteins the finger most closely resembles the carboxy terminal finger.

1.1.2.3 The Cys_6 binuclear cluster

Another type of finger, the Cys_6 binuclear cluster (Cys-X_2-Cys-X_6-Cys-X_{5-16}-Cys-X_2-Cys-X_{6-8}-Cys), has to date only been found in fungal regulatory proteins, for example the *A. nidulans* proteins AMDR (Andrianopoulos and Hynes, 1990), ALCR (Kulmburg *et al.*, 1991) and UAY (Suarez *et al.*, 1995). This class is exemplified by the *S. cerevisiae* GAL4 protein (Pan and Coleman, 1990). GAL4 is a positive regulatory protein involved in the utilisation of galactose. GAL4 binds as a dimer to the 5' region of genes with a recognition sequence that has approximate dyad symmetry (Marmorstein *et al.*, 1992 and references therein). X-Ray crystallographic studies have shown that within the first 65 amino acids there is a DNA recognition domain (8-40 amino acids), a linker (41-49) and a dimerisation domain (49-65) (Marmorstein *et al.*, 1992). Within the 17 bp binding site is a highly conserved CCG sequence

at either end. The two recognition helices lie in the major groove centred over the conserved CCG triplet and are stabilised by interactions with the phosphate backbone. The triplets are separated by about one to one and a half turns of the DNA helix. This position puts the carboxy end of the α helix into the major groove so that Lys17 and Lys18 make specific contacts with the CCG base sequence. There are no other contacts between amino acid side chains and bases and it is not possible to change the specificity of this type of DNA binding protein by altering the side chains at this position. Other yeast proteins containing a Cys_6 binuclear cluster, LAC9 and PPR1, also recognise this CCG triplet and have a Lys18 in an equivalent position (Marmorstein *et al.*, 1992 and references therein).

1.1.2.4 The Cys_8 zinc cluster

The glucocorticoid receptor is part of the nuclear receptor superfamily which includes receptors for the steroid hormones and certain regulators of development in *Drosophila*. They share a DNA binding domain characterised by eight cysteines coordinated by two Zn^{2+} ions in a binuclear cluster (Luisi *et al.*, 1991 and references therein). The TFIIIA zinc fingers act as independent structures able to bind DNA whereas the zinc cluster folds together as part of a larger globular domain in order to interact with DNA.

1.1.2.5 The RING-finger motif

A possible member of the zinc containing DNA binding motif family may be the RING-finger motif. This motif is characterised by the sequence Cys-X-(I/L/V)-Cys-X_{10-30}-Cys-X-His-X-(F/I/L)-Cys-X_2-Cys-(I/L/M)-X_{9-18}-Cys-P-X-Cys where the first and third pairs of Cysteine residues co-ordinate one zinc cation and the second and fourth pairs of cysteine/ histidine residues co-ordinate another zinc cation (Ishida *et al.*, 1993; Everett, *et al.*, 1993). Only weak binding to DNA cellulose columns has so far been demonstrated for proteins containing this motif (Tagawa *et al.*, 1990; Lovering *et al.*, 1993), and Everett *et al.*, (1993) failed to find stable interactions with DNA using an expressed polypeptide. However this motif is present within a protein which contains motifs suggestive of a role in transcriptional regulation (Tranque, *et al.*, 1996).

1.1.3 The Leucine zipper and helix loop helix motifs

The motifs mentioned above all participate in DNA binding, however, there are other motifs which are strictly speaking involved in protein-protein interactions that are required before DNA binding can occur. These are the leucine zipper (Landschulz *et al.*, 1988) and the related basic helix-loop-helix motif (bHLH) (Murre *et al.*, 1989). In the leucine zipper motif leucine residues occur every seventh amino acid in an helical structure resulting in leucines on one face of the helix on every second turn. Oppositely charged amino acids are juxtaposed in a manner suitable for intrahelical ion pairing and form a very stable helix (reviewed by Johnson and McKnight, 1989). The leucine zipper facilitates dimerisation of two proteins containing this motif and the basic regions adjoining the leucine zippers are able to make the correct contacts with the DNA helix (reviewed by Busch and Sassone-Corsi, 1990). Examples of proteins with leucine zippers and basic regions adjoining them are GCN4 from *S. cerevisiae*, the proto oncogenes Fos and Jun and the mammalian CCAAT/enhancer binding protein C/EBP (Landschulz *et al.*, 1988). Like the leucine zipper motif the basic HLH (bHLH) motif allows dimerisation (through one of the helices) and binding by the basic region. This motif is found in the muscle determination factor MyoD and the immunoglobulin κ-chain enhancer binding proteins E12 and E47 (reviewed by Baxevanis and Vinson, 1993). The oncogene products Myc and Max contain both a bHLH and leucine zipper motif (Prendergast and Ziff, 1992 and references therein). Recently the LIM domain, a cysteine rich motif present in a diverse array of proteins, has been characterised (reviewed by Sanchez-Garcia and Rabbitts, 1994). This motif binds zinc much like zinc finger proteins but is involved in protein-protein interactions (Gong and Hew, 1994; Schmeichel and Beckerle, 1994).

1.1.4 Other DNA binding motifs

Despite their differences, the DNA binding motifs mentioned all result in the placing of an α-helix into the major groove of DNA. There are some motifs being studied that result in base specific contacts of a β-sheet with DNA, for example the Arc repressor from bacteriophage P22 (Raumann *et al.*, 1994). In addition some proteins bind with unique binding motifs. For example the crystal structure of the p53 protein shows a new motif that is functionally

inactivated by point mutations not just within amino acids that contact the DNA but also in those that have a structural role. In addition, P53 recognises the minor groove of DNA as opposed to the major groove (Cho *et al.*, 1994). A finger structure with three cysteines and one histidine has been found in a retrovirus and binds DNA in a very similar fashion to other zinc fingers (Summers *et al.*, 1990).

1.2 Carbon Catabolite Repression

Carbon catabolite repression has been shown to occur in single cell prokaryotes, for example *E. coli* (section 1.2.1), and in simple eukaryotes like the yeast *S. cerevisiae* (section 1.2.2) and the filamentous fungus *A. nidulans* (section 1.2.3). Carbon sources differ in their ability to bring about carbon catabolite repression and this difference can be species specific. For example, within the genus *Aspergillus*, glycerol is a moderately derepressing carbon source for *A. nidulans* but a strongly repressing one for *A. niger* (O'Connell, 1990). How these differences occur is not known as little is understood about the signal produced in response to glucose even in the well studied prokaryote *E. coli*. However, the molecular machinery involved in bringing about a change in gene expression of the lactose operon of *E. coli* is well characterised and is presented below.

1.2.1 Carbon Catabolite Repression in *E. coli*

When *E. coli* cells are grown in the presence of two carbon sources, for example, glucose and lactose, they exhibit two step diauxic growth. That is, the utilisation of lactose does not occur until all of the glucose in the medium is gone. The presence of glucose causes the repression of the enzymes required for the utilisation of lactose. This was first called the "glucose effect" but later, other carbon sources were also found to cause this repression and so it is now referred to as carbon catabolite repression (Pastan and Perlman, 1970, and references therein; Magasanik, 1961).

The presence of glucose causes at least three major changes within the cell (reviewed by Kolb *et al.*, 1993). Firstly, glucose inhibits the uptake of inducer molecules necessary for the expression of genes involved in the utilisation of alternative carbon sources (inducer exclusion). Secondly, glucose causes a transient repression of enzyme activity (catabolite inactivation) and thirdly glucose lowers the intracellular concentration of cAMP which acts to repress enzyme synthesis.

The cAMP level in the cell is under very complex control but perhaps the most important point of control is the activity of the enzyme adenylate cyclase which catalyses the formation of cAMP from ATP. Adenylate cyclase deficient mutants are defective in carbon catabolite repression and this defect is overcome by the exogenous addition of cAMP (Pastan and Perlman, 1970). The principal control of adenylate cyclase is post-transcriptional via the phosphotransferase system (PTS) (reviewed by Postma *et al.*, 1993). Another gene identified as being central to carbon catabolite repression is the *crp* (the cAMP receptor protein) gene. The CRP protein is a transcriptional activator required for the expression of many genes including those under carbon catabolite repression (reviewed by Kolb *et al.*, 1993), starvation control (Schultz *et al.*, 1988) and anaerobiosis (Jerlstrom *et al.*, 1987). Recent work by Ishizuka *et al.*, (1993, 1994) has shown that *crp* transcription is down regulated by a process involving CRP and cAMP in the presence of glucose.

The lactose operon of *E. coli* is regulated by catabolite repression. Maximal gene expression only occurs when glucose becomes limiting. In the absence of glucose, the high level of cAMP within the cell encourages the formation of the cAMP-CRP complex. The CRP protein exists as a dimer with each subunit binding one molecule of cAMP, although the most active form of this complex contains only one bound cAMP moiety. The CRP protein contains a HTH DNA binding motif (section 1.1.1) and a cAMP binding domain (Weber and Steitz, 1987 and references therein). The cAMP-CRP complex is able to bind to the 5' region of the lactose operon in close proximity to the RNA polymerase promoter and increase the rate of

transcription. Although the precise mechanism is not yet known the simplest explanation is that CRP contains a region which is able to directly interact with RNA polymerase to increase transcription initiation (Kolb *et al.*, 1993). Alternatively the observation that CRP-cAMP bends DNA when it binds suggests that perhaps CRP-cAMP increases transcription by making the promoter region more "attractive" to RNA polymerase.

1.2.2 Carbon Catabolite Repression in *S. cerevisiae*

Carbon catabolite repression in yeast has been extensively studied in the species *S. cerevisiae*. As reviewed by Johnston and Carlson (1992), numerous genes have been shown to be under carbon catabolite repression including: genes for the utilisation of carbon sources that are alternatives to glucose, such as galactose (*GAL*), maltose (*MAL*) and sucrose (*SUC*); genes that are not required at high levels during fermentation, for example enzymes of the TCA cycle, glyoxylate by-pass, and mitochondrial proteins in the respiratory chain; and enzymes responsible for gluconeogenesis, sporulation, glucose transport, proteases, peroxisomal enzymes and acetyl CoA hydrolase. In all the cases studied researchers have demonstrated control at the level of transcription.

The degree to which the presence of glucose exerts repression depends on the gene involved. For example invertase (responsible for the extracellular hydrolysis of sucrose) can be repressed 750-fold by glucose (Estruch and Carlson, 1990a), whereas aconitase is repressed only 5-fold (Polakis and Bartley, 1965). A complicating factor in studies of the yeast *S. cerevisiae* has been the variation in the degree of carbon catabolite repression for the same gene between different strains, and the variation in the number, location and structure of genes between strains (reviewed by Carlson, 1987).

A number of mechanisms, in addition to carbon catabolite repression, have been found in *S. cerevisiae* that lead to decreased enzyme activity in the presence of glucose. These include inducer exclusion, catabolite inactivation, and decreased mRNA stability, although not all

mechanisms operate on every gene studied. The rapid inactivation of certain enzymes by the presence of glucose is termed catabolite inactivation and has been shown to occur in *S. cerevisiae* and some other yeasts (reviewed by Holzer, 1976, Wills, 1990). For the gluconeogenic enzyme fructose-1,6-bisphosphatase, the first step is thought to be a reversible inhibition of the enzyme by phosphorylation (Mazon *et al.*, 1982a, b) followed by irreversible proteolytic degradation by the proteosome, a multisubunit enzyme complex harbouring several different proteolytic activities (Schork *et al.*, 1994 and references therein). Proteolytic cleavage is independent of the phosphorylation state of the protein (Rose *et al.*, 1988). Researchers have also found a marked decrease in the stability of the mRNA of certain enzymes (eg. *MAL* genes: Federoff *et al.*, 1983). The literature reviewed in this section will focus on the factors which operate at the level of the initiation of transcription.

It is not yet known what signal is produced by cells in response to glucose. However the presence of a functional kinase (hexokinase PI, PII or glucokinase) is necessary (Beullens *et al.*, 1988). It seems unlikely that cAMP plays the same central role in yeast as it does in *E. coli,* as Matsumoto *et al.*, (1983) have studied mutants unable to synthesise cAMP and found that synthesis of galactokinase, which is sensitive to catabolite repression, was not affected by the levels of cAMP in the cell. This is contrary to the situation in *E. coli* where mutants lacking adenylate cyclase are unable to utilise carbon sources such as lactose and maltose (Pastan and Perlman, 1970). Other researchers have found that the level of intracellular cAMP rises transiently in response to glucose (cAMP spike), and suggest that it is this which starts a cascade of phosphorylation /dephosphorylation events which activate phosphofructokinase and inactivate 1,6-diphosphatase (Beullens *et al.*, 1988).

Mutational analysis has not identified a single controlling gene for carbon catabolite repression in yeast but instead a system of degenerate pathways with separate genes required for repression and for derepression of regulated genes. Mutations affecting carbon catabolite repression have been isolated in a number of laboratories and so the same gene has been given many names. A

unified agreement on nomenclature has not yet been reached but in this thesis a commonly accepted name has been used and alternative names can be found in Table 1.1. An overall summary of how some of these regulatory proteins may work together to bring about regulated expression of the *SUC2* gene is presented in section 1.2.2.1. A description of some of the genes involved in derepression and repression is then given to draw the readers attention to points of interest especially where relevant to the work presented in this study.

1.2.2.1 Regulation of the *SUC2* gene

The *SUC2* gene of *S. cerevisiae* encodes a secreted invertase and is the subject of much investigation into the molecular mechanism of carbon catabolite repression. Since *SUC2* is not induced by a carbon source, its regulation was presumed to be less complex than for other genes of the *GAL* and *MAL* pathways. However, many genes are required for maximal derepression and many are required for repression to occur (reviewed by Gancedo, 1992; Johnston and Carlson, 1992). It is beyond the discussion presented here to mention all the genes which affect *SUC2* expression, and for many little is known about their function. Control of *SUC2* expression includes global repression and global activation elements as well as a myriad of specific *trans* acting factors. A simple model for the control of *SUC2* expression is summarised in Figure 1.2 (Gancedo, 1992).

MIG1 is a DNA binding protein able to bind to the 5' region of *SUC2.* In the presence of glucose mutations in the *MIG1* gene result in derepressed levels of invertase. The MIG1 protein does not directly repress *SUC2* transcription initiation but recruits the SSN6-TUP1 repressor complex to carry out the repression (Nehlin and Ronne, 1990; Keleher *et al.*, 1992; Treitel and Carlson, 1995). MIG1 is differentially phosphorylated in response to glucose availability (Treitel and Carlson, 1995).

SNF1 mutants fail to derepress *SUC2* at the level of transcription and genetic evidence suggests that one of the functions of SNF1 is to relieve transcriptional repression mediated by the SSN6-

Table 1.1 Alternative names for some of the genes involved in carbon catabolite repression in the yeast *S. cerevisiae*[a]

Genes required for derepression	Function (if known)
SNF1=CAT1=CCR1	SNF1 is a Serine/ threonine protein kinase and SNF4 may be a positive activator of SNF1
SNF4=CAT3	
HAP2	CCAAT binding transcriptional activator complex
HAP3	
HAP4	
HAP5[b]	
SNF2=GAM1=SWI2=TYE3	transcriptional activator complex involved in relieving transcriptional repression by chromatin.
SNF5=TYE4=SWI10[c]	
SNF6	
SWI1[c]*=ADR6=GAM3*[d]	
SWI3[c]*=TYE2*[e]	
MSN1=FUP1[f]	activator of transcription
MSN2[g]	?

Genes required for repression	Function (if known)
HXK2=HEX1=GLR1	structural gene for hexokinase PII
HEX2=REG1	negative regulatory element?
TUP1=FLK1=UMR7=CYC9=AMM1=SLF2= AER2=AAR1	transcriptional repressor complex
SSN6=CYC8	
MIG1=CAT4[h]*=SSN1*[i]	zinc finger DNA binding protein
RGR1	?
CAT80=GRR1	required for intracellular signal?
SPT6=SSN20=CRE2[j]	?

a Gancedo, (1992) and Johnston and Carlson, (1992)
b McNabb *et al.*, (1995)
c Winston and Carlson, (1992)
d Peterson *et al.*, (1994)
e Lohning *et al.*, (1993)
f Estruch and Carlson, (1990b)
g Estruch and Carlson, (1993)
h Entian and Barnett, (1992)
i Vallier and Carlson, (1994)
j Swanson *et al.*, (1990)

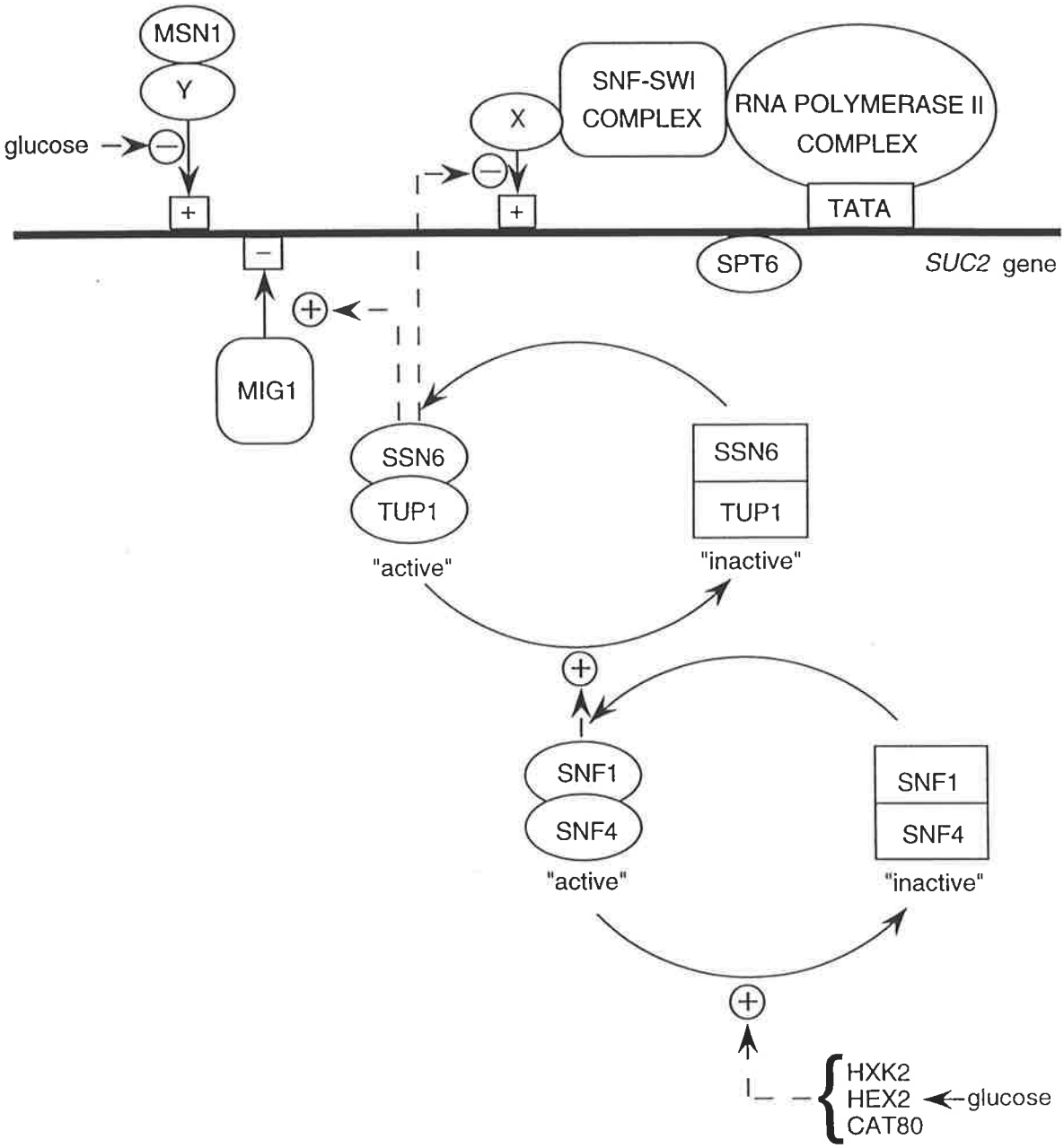
TUP1 repressor complex (Carlson and Botstein, 1982; Keleher *et al.*, 1992; Schultz and Carlson, 1987). The *SNF1* gene encodes a serine/threonine protein kinase which may act to modify SSN6-TUP1 function, although SSN6 is probably not a direct target for the SNF1 protein kinase (Celenza and Carlson, 1986; Schultz *et al.*, 1990). SNF1 is physically associated with a positive regulator, SNF4, as well as other proteins (Fields and Song, 1989; Celenza *et al.*, 1989; see also section 1.2.2.2a). The signal produced by cells in response to glucose is not known but the products of the *HXK2*, *HEX2* and *CAT80* genes are likely to be involved in signalling the presence of glucose to the SNF1-SNF4 protein complex.

The products of the *SWI1, SWI3, SNF2, SNF5*, and *SNF6* genes form a large multimeric complex which plays a general role in the activation of transcription of many genes (Peterson *et al.*, 1994 and references therein). These proteins are thought to associate with gene specific activators to relieve transcriptional repression by chromatin (reviewed by Roth, 1995). Since the complex lacks a DNA-binding motif perhaps it is acting as an adaptor (Berger *et al.*, 1990) or a mediator (Flanagan *et al.*, 1991) between the RNA polymerase II complex and an as yet unidentified *SUC2* specific DNA-binding protein X.

Disruption of the *MIG1* gene results in a significant but partial relief of glucose repression of *SUC2* indicating that there are MIG1 independent pathways of repression (Nehlin and Ronne, 1990). For example, the SSN6-TUP1 complex might interfere with binding and/or activity of protein X. A strain carrying a disrupted *MSN1* gene has a few fold lower level of *SUC2* expression in derepressed conditions (Estruch and Carlson, 1990b). A lexA-MSN1 fusion protein is able to activate transcription but MSN1 has only a weak non specific DNA-binding activity. Thus, if it does act as an activator it probably does so in conjunction with other proteins (denoted as "Y" in Fig. 1.2; Estruch and Carlson, 1990b). MSN1 is not absolutely required for *SUC2* expression but when present in multiple copies it overcomes the requirement for SNF1 (Estruch and Carlson, 1990b). Additional pathways other than through SNF1 and SSN6-TUP1 must exist since a *snf1* mutant with multiple copies of the *MSN1* gene, is still able

Figure 1.2 A simplified scheme summarising the regulated expression of the *SUC2* gene

In the absence of glucose, the positively acting pathways working through the regulatory proteins MSN1 and X result in transcription of the *SUC2* gene. In the presence of glucose the HXK2, HEX2 and CAT80 proteins favour the formation of an inactive SNF1/SNF4 complex which is not able to inactivate the SSN6/TUP1 repressor. The MIG1 protein recruits the SSN6/TUP1 repressor to repress *SUC2* transcription. In addition, the SSN6/TUP1 repressor may inhibit the activator protein X. Glucose causes the down regulation of MSN1 via a pathway independent of SNF1. For a more detailed description of the regulatory proteins mentioned here (including SPT6) see text. Shown is a modification of Figure 3 from Gancedo, (1992).


MSN1
Y
glucose
X
SNF-SWI
COMPLEX
RNA POLYMERASE II
COMPLEX
TATA
SPT6
SUC2 gene
MIG1
SSN6
TUP1
"active"
SSN6
TUP1
"inactive"
SNF1
SNF4
"active"
SNF1
SNF4
"inactive"
HXK2
HEX2
CAT80
glucose

to be partially repressed. Other DNA binding proteins MSN2 and MSN4 have been found to have a slight positive effect on *SUC2* expression but it is not known whether they achieve this via a direct or indirect mechanism. Perhaps they can compete with MIG1 for binding since they both encode zinc finger proteins with high sequence similarity to MIG1 (Estruch and Carlson, 1993).

Although it is unlikely that cAMP plays a central role, there are various lines of evidence to suggest that cAMP plays some role in the control of *SUC2* expression. However, the nature of the involvement of cAMP in the molecular mechanism of carbon catabolite repression remains unclear. SKO1, which appears to be turned on by cAMP-dependent protein kinase, binds to the 5' region of *SUC2* and acts negatively (Nehlin *et al.*, 1992). *sko1* is not a suppressor of *snf1,* therefore SKO1 acts in a different pathway to MIG1. The *rgr1* mutation was isolated as allowing overexpression of a gene under control of the *SUC2* promoter. This mutation results in elevated cAMP and CDC25 levels (Sakai *et al.*, 1990). Overexpression of the *CDC25* gene causes derepression (Van Aelst *et al.*, 1991), and thus perhaps it is this rather than the altered cAMP levels that is the cause of the derepression observed. However, as CDC25 is involved in signalling the presence of glucose to adenylate cyclase by the RAS1 and RAS2 proteins (Van Aelst *et al.*, 1991), the involvement of cAMP in derepression seen in *rgr1* mutants is not clear. Cells containing a mutation in the *HEX2* gene show continuous glucose triggered activation of cAMP synthesis. The glucose triggered cAMP spike is responsive to a glucose repressible component which turns off the glucose induced increase in cAMP, and allows proper entry into the stationary phase (Go) (Dumortier *et al.*, 1995). CDC25 is not, however, the glucose responsive component of the signalling pathway (Van Aelst *et al.*, 1991).

The rate of *SUC2* transcription depends on a balance between activation by MSN1 and protein X (together with SWI1/3/SNF2/5/6) and inhibition by MIG1 and other MIG1 independent systems. The possibility that glucose affects the transcription rates of the various *trans* acting factors seems unlikely. The steady-state levels of the mRNAs have been studied in most cases

and found not to be dependent on the presence or absence of glucose (see Gancedo, 1992 and references therein). Therefore it seems that glucose acts by modifying the activity of these regulatory proteins.

1.2.2.2 Genes found to be required for derepression:

a) *SNF1, SNF4*: The importance of the *SNF1* (Sucrose non-fermenting) gene is indicated by the inability of *snf1* mutants to grow on a wide range of fermentable and non-fermentable carbon sources (Celenza and Carlson, 1984 and references therein; Entian and Zimmermann, 1982).The SNF1 protein contains a serine/threonine protein kinase domain which is essential for its function, and SNF1 is distributed throughout the cell (Celenza and Carlson, 1986). Genetic evidence suggests that one of the functions of SNF1 is to relieve transcriptional repression mediated by the SSN6-TUP1 repressor complex (section 1.2.2.3c). The SNF1 protein has been found to interact with the SNF4 protein via protein-protein interactions (Fields and Song, 1989; Celenza *et al.*, 1989). SNF4 is not a substrate of SNF1 but is thought to be a positive regulator of SNF1. A role in glucose signalling has not been postulated for SNF4 (Celenza and Carlson, 1989, Celenza *et al.*, 1989). Other proteins have been found to interact with SNF1 in a two-hybrid screen, including SIP1, SIP2, SIP3, SIP4 and MSN3 (Yang *et al.*, 1992; Albert Hubbard *et al.*, 1994). SIP1 and SIP2 share sequence similarity with each other and another protein GAL83 (Erickson and Johnston, 1993; Yang *et al.*, 1994). High copy numbers of the SIP1, SIP2 and HEX2 genes are able to complement a *gal83* mutation and thus all are probably functionally related (Erickson and Johnston, 1993). It is suggested that these proteins act as adaptors between SNF1 and specific targets of the SNF1 kinase (Yang *et al.*, 1994). The SIP3 protein contains a leucine zipper motif and is able to function as a transcriptional activator (Lesage *et al.*, 1994). SIP4 contains a Cys_6 binuclear cluster DNA binding motif and at least one motif involved in protein-protein interactions (Lesage *et al.*, 1996). Transcriptional activation by SIP4 is regulated by glucose and is dependent on SNF1, however, it is not yet known which genes SIP4 regulates (Lesage *et al.*, 1996). SIP4 is differentially phosphorylated in response to glucose and this phosphorylation is dependent on SNF1 (Lesage *et al.*, 1996).

A *snf1* mutant was complemented by a cDNA clone from rye endosperm suggesting that the regulatory mechanism used for yeast carbon catabolite repression is conserved among eukaryotes (Alderson *et al.*, 1991). A mammalian homologue of SNF1, AMP activated protein kinase (AMPK), has been identified (Carling *et al.*, 1994). In addition to the structural similarity of these two proteins, their role in the regulation of the activity of at least one enzyme, acetyl-CoA carboxylase, has been highly conserved (Woods *et al.*, 1994). SNF1 activity increases markedly during glucose-derepressing conditions and this is a result of reversible phosphorylation (Woods *et al.*, 1994).

b) *HAP2, HAP3, HAP4, HAP5*: The *hap2, hap3* and *hap4* mutants are unable to grow on non-fermentable carbon sources like lactate. These genes are positive regulators required for the expression of cytochromes and other proteins required during oxidative phosphorylation. The HAP2, HAP3 and HAP5 proteins form a heterotrimeric DNA binding complex able to bind a CCAAT recognition sequence and associate in the absence of DNA (Hahn and Guarente, 1988; Xing *et al.*, 1993, McNabb *et al.*, 1995). CCAAT sequences are recognised by many regulatory proteins with specificity dependent on flanking sequences. HAP2 and HAP3 are functionally conserved since the mammalian CCAAT-binding factor can substitute for binding at the CCAAT box (Chodosh *et al.*, 1988). HAP4 contains an acidic domain which when associated with HAP2 and HAP3 provides the complex with an activation domain. All four proteins are required for activation at the CCAAT box. The HAP2/3/4 complex has been conserved among yeasts and mammals, although its cellular function has been altered during evolution (Becker *et al.*, 1991; Olesen *et al.*, 1991). *HAP4* (Forsburg and Guarente, 1989) and *HAP2* (Pinkham and Guarente, 1985) are induced at the level of mRNA by a change in carbon source from glucose to a non-fermentable carbon source is likely to account for the regulation of the many genes under HAP2/3/4 control. This is likely to involve a relief of carbon catabolite repression since at least one of the genes under HAP2/3/4/5 control, iso-1-cytochrome *c* (CYC1), is derepressed in the presence of a mutation in MIG1 (Schuller and Entian, 1991).

c) *SWI1, SWI3, SNF2, SNF5, SNF6*: The *SWI1*, *SWI3*, *SNF2, SNF5* and *SNF6* gene products are required for the transcription of many divergently regulated genes (Johnston and Carlson, 1992 and references therein). It appears that these proteins form a large multiprotein complex that associates with gene specific proteins to assist in relieving transcriptional repression by chromatin (Cairns *et al.*, 1994; Peterson *et al.*, 1994 and references therein). Biochemical studies have found altered chromatin structure of the *SUC2* promoter in *snf2* and *snf5* mutants and this alteration is suppressed by a mutation in a histone encoding gene (Hirschhorn *et al.*, 1992). The SNF2/5/6 proteins are not required in a *spt6* background or when large regions of the promoter have been removed (Neigeborn *et al.*, 1987). *SPT6* is an essential gene possibly with a general role in transcription or perhaps in chromatin structure (Swanson *et al.*, 1990). At least five additional peptides have been implicated as components of the large multiprotein SNF-SWI complex. One of these, SNF11 has recently been identified and shown to interact with a region of SNF2 (Treich *et al.*, 1995). SNF2 has been shown to have an ATPase activity and contains a region with sequence similarity to helicases (Laurent *et al.*, 1993; Davis *et al.*, 1992). The SWI3 protein has been shown to associate with a gene specific activator and this association was dependent on the presence of other proteins within the complex (Yoshinaga *et al.*, 1992). The SNF5 protein is a glutamine and proline rich transcriptional activator able to affect expression of a wide range of genes (Laurent *et al.*, 1990). The isolation of SNF2 and SNF5 homologues form *Drosophila* and humans suggests functional conservation of this regulatory mechanism in higher eukaryotes (reviewed by Carlson and Laurent, 1994).

1.2.2.3 Genes found to be required for repression:

The genes required for repression are especially relevant because mutations in these genes have the same phenotype as the *creA* mutants in *A. nidulans* (section 1.2.3.2) namely that of derepression.

a) *HXK2*: Hexokinase PII (HXK2) is the main isozyme of hexokinase responsible for glucose repression. HXK2 mutants have derepressed levels of many genes in the presence of glucose

(Entian, 1980). The catalytic activity of hexokinase PII is directly related to the level of carbon catabolite repression seen as measured by invertase activity (Ma *et al.*, 1989; Rose *et al.*, 1991). The minor hexokinase PI (HXK1), but not glucokinase, was able to mediate glucose repression when overproduced (Rose *et al.*, 1991). Phosphorylation is thus a significant controlling factor of regulation. This may be a function of the amount of autophosphorylation of the enzyme hexokinase PII, as suggested by the results of Vojtek and Fraenkel (1990). Alternatively regulation may be related to the phosphorylation of some other substrate which could be metabolic or regulatory.

b) *HEX2*: The *hex2* mutation causes pleiotrophic effects including increased hexokinase PII activity, inhibited growth on maltose, and glucose insensitive expression of invertase (Entian, 1981 and references therein). The HEX2 amino acid sequence does not contain any features that indicate function, however, HEX2 does appear to be localised in the nucleus (Niederacher and Entian, 1991). HEX2 is associated with a protein phosphatase type I (PP1 or GLC7), and may be a regulatory subunit of it. The phosphatase targets proteins in the carbon catabolite repression pathway (Tu and Carlson, 1995). It is thought that HEX2 controls the activity of PP1 toward substrates that are phosphorylated by SNF1 (Frederick and Tatchell, 1996).

c) *SSN6,TUP1*: The *ssn6* and *tup1* mutations cause pleiotrophic defects including derepression of glucose repressible genes, non-sporulation of homozygous diploids and mating type defects in MATα cells, suggesting an involvement in many regulatory pathways (Johnston and Carlson, 1992 and references therein). The SSN6 and TUP1 proteins are associated together and act as a general repressor of transcription being recruited to repress specific genes by DNA binding proteins like **a**1, α2, MCM1, ROX1 and MIG1 (Williams *et al.*, 1991, Keleher *et al.*, 1992; Treitel and Carlson, 1995; Herschbach *et al.*, 1994). TUP1 contains a WD repeat (first identified in β-transducin) which directly interacts with α2 (Williams and Trumbly, 1990; Komachi *et al.*, 1994). SSN6 contains ten copies of a tetratricopeptide (TPR) motif, some of these have been found to be responsible for interactions with TUP1, others are required for repression of oxygen

regulated genes (via ROX1) and yet others are required for the repression of glucose repressed genes (via MIG1; Sikorski *et al.*, 1990; Tzamarias and Struhl, 1995). It is thought that these different motifs mediate protein-protein interactions linking the SSN6-TUP1 complex to structurally dissimilar DNA binding proteins required for pathway specific regulation, and that repression occurs via the TUP1 repression domain in each case (Tzamarias and Struhl, 1994; 1995).

d) *MIG1*: *MIG1* (Multicopy Inhibitor of *GAL* gene expression) was cloned as inhibiting expression of a lethal gene under the control of the *GAL1* promoter (Nehlin and Ronne, 1990). *MIG1* was found to be allelic to *CAT4* and *cat4* was isolated as a suppressor of *snf1* and *snf4* by restoring growth on raffinose (Schuller and Entian, 1991). The MIG1 protein contains two zinc finger DNA-binding motifs of the Cys_2His_2 type with sequence similarity to the zinc fingers of the mammalian Zif268 protein (section 1.1.2.1), WT1 (Call *et al.*, 1990) and the putative zinc fingers of the CREA protein from *A. nidulans* (section 1.2.3.3). The MIG1 protein binds to the *SUC2*, *GAL1* and *GAL4* promoters with a consensus sequence of A/T A/T A/T A/T T G/C C/T GGGG. A saturated mutagenesis of one of the naturally occurring *SUC2* sites has recently been carried out by Lundin *et al.*, (1994) and will be discussed in section 3.8.1. As mentioned previously MIG1 recruits the SSN6-TUP1 repressor to carry out the repression. Recent work has suggested that perhaps SNF1 is able to directly inactivate MIG1 (section 6.6.2; Ostling *et al.*, 1996) although a role for SNF1 in inactivating SSN6-TUP1 is not excluded. SSN6 masks a positive domain on MIG1, indicating that perhaps MIG1 can act as an activator or a repressor depending on the circumstances, however, the significance of this is unknown (Treitel and Carlson, 1995). MIG1 represses the transcription of the structural genes *SUC2* and *GAL1*, as well as the regulatory gene for *GAL1*, *GAL4* (Nehlin and Ronne, 1990; Nehlin *et al.*, 1991). There is also evidence to suggest that MIG1 is involved in down regulating the induction pathway of *GAL* genes (Nehlin *et al.*, 1991) and thus MIG1 is implicated in all the steps leading to glucose repression of the *GAL* genes (Johnston *et al.*, 1994).

e) *RGR1*: The *rgr1* mutation was isolated as allowing overexpression of a gene under the control of the *SUC2* promoter. The *rgr1* mutant showed pleiotrophic effects for the following cellular functions: (1) resistance to carbon catabolite repression, (2) temperature sensitive lethality, (3) sporulation deficiency in homozygous diploids and (4) abnormal cell morphology (Sakai *et al.*, 1988). *RGR1* has been cloned and found to encode a 123 kDa protein with several membrane spanning regions but no other sequence similarity to other proteins in the data base (Sakai *et al.*, 1990). A null mutant is lethal and abnormal mutant morphology may be due to reduced levels of stored carbohydrates since Sakai *et al.*, (1990) have observed a higher level of intracellular cAMP in *rgr1* mutants compared to wild type. The enzymes catalysing the degradation of reserve carbohydrates, glycogen phosphorylase and trehalase, are activated by cAMP-dependent phosphorylation (Uno *et al.*, 1983; Wingender-Drisson and Becker, 1983). RNA polymerase II from *S. cerevisiae* exists in holoenzyme forms containing a complex, commonly referred to as the mediator, associated with its carboxyl-terminal domain. Certain SSN proteins, SIN4 and RGR1 are now thought to associate with the mediator complex or may in fact be a part of it (Li *et al.*, 1995; Song *et al.*, 1996). SIN4, GAL11 and at least one other peptide are dependent on the presence of a functional RGR1 protein for this association (Li *et al.*, 1995). Given that RGR1 acts negatively on *SUC2* expression this work suggests that the mediator complex is involved in both transcriptional repression and activation (Li *et al.*, 1995). A role for RGR1 in affecting the structure of chromatin has been implicated in studies of RME1-dependent repression (Covitz, *et al.*, 1994), however it is not clear how to integrate these results into a role for the RGR1 protein.

1.2.2.4 Summary

It is clear that a multitude of genes act concertedly to bring about the repression and derepression of genes under carbon catabolite control. Even an enzyme such as invertase which is not induced by its substrate has a complex array of *trans* acting factors controlling regulated expression. Post-translational control via phosphorylation/dephosphorylation is a likely to play an important role in carbon catabolite repression in *S. cerevisiae*.

1.2.3 Carbon Catabolite Repression in *A. nidulans*

In *A. nidulans*, as is the case with *E. coli* and *S. cerevisiae*, the presence of a source of carbon catabolite repression, such as glucose, results in decreased enzyme activities for enzymes of the glyoxylate by-pass, respiratory enzymes and the enzymes and permeases required for the utilisation of novel carbon sources (as detailed in Arst and Bailey, 1977)

Early work does not suggest a central role for cAMP in the carbon catabolite repression response of *A. nidulans* (discussed by Arst and Bailey, 1977 and references therein) as neither the exogenous addition of adenylate cyclase activators, such as glucagon and isoproterenol, nor inhibitors of cAMP phosphodiesterases such as theophylline and aminophylline, alter the levels of enzymes which are subject to carbon catabolite repression. Unlike the situation in *S. cerevisiae*, catabolite inactivation does not appear to play a role in *A. nidulans*. A number of enzymes have been found to be unaffected by the presence of glucose or sucrose (Kelly, 1980). This has allowed a more straight forward interpretation of the mutants affected in carbon catabolite repression. An exception to this is the rapid inactivation of isocitrate lyase by glucose via a process not involving *creA* (De Lucas *et al.*, 1994a).

An investigation of carbon catabolite repression has focused on the control of transcription of regulated genes as it has been shown that *creA* affects the level of transcription of a number of gene systems. These include genes required for the utilisation of alternative carbon sources, like acetamide *amdS* (Hynes *et al.*, 1983), ethanol, *alcA* and *aldA* (Lockington *et al.*, 1985, 1987), lactams *lamA* and *lamB* (Katz and Hynes, 1989a), acetate, *facA* and *acuE* (Sandeman and Hynes, 1989) and proline *prnB* (Sophianopoulou *et al.*, 1993) as well as affecting transcription of regulatory genes, for example *facB* (Katz and Hynes, 1989b) and *alcR* (Lockington *et al.*, 1985, 1987). This section describes the isolation of *creA* mutants and subsequent cloning and preliminary characterisation of the *creA* gene.

1.2.3.1 Selection of mutants affected in carbon catabolite repression

Mutants defective in carbon catabolite repression were selected in a number of ways. The two most commonly used were as suppressors of *areA* and *pdhA* alleles.

The *areA* gene codes for a positively acting regulatory protein involved in nitrogen metabolite repression. Certain *areA* mutants are unable to utilise any nitrogen source other than ammonium because they fail to derepress the enzymes required for the utilisation of nitrogen sources that are alternatives to ammonium, even in the absence of a source of nitrogen metabolite repression. Substrates which provide both a carbon and a nitrogen source (for example acetamide or L-proline) can be utilised in the absence of a source of carbon catabolite repression (Arst and Cove, 1973). Thus mutants at the *creA* locus were isolated that allowed *areA*$^-$ strains to utilise acetamide and L-proline as nitrogen sources in the presence of glucose and sucrose (Arst and Cove, 1973; Hynes and Kelly, 1977)

The other method used for the selection of *creA* alleles involved the suppression of mutants lacking a functional pyruvate dehydrogenase (*pdhA*) gene. *pdhA*$^-$ strains are unable to produce acetyl CoA from pyruvate (Romano and Kornberg, 1968, 1969). Substrates such as acetamide and ethanol can provide the cell with a source of acetyl CoA provided that the genes encoding the enzymes needed for the conversion of these to acetate are not subjected to carbon catabolite repression (Bailey and Arst, 1975). Therefore *creA* mutants were selected as allowing *pdhA*$^-$ strains to utilise ethanol or acetamide to provide a source of acetyl CoA in the presence of glucose (Bailey and Arst, 1975; Arst and Bailey, 1977). The degree to which carbon sources prevent *areA*$^-$ strains from utilising L-proline, acetamide and GABA closely resembles the degree to which they prevent ethanol supplementation of *pdhA*$^-$ strains (Arst and Bailey, 1977). Thus both systems measure the amount of carbon catabolite repression exerted by a carbon source.

1.2.3.2 Phenotype and physical characterisation of mutants affected in carbon catabolite repression

Despite the existence of strong selection procedures, for example resistance to fluoroacetate and fluoroacetamide (toxic analogues of acetate and acetamide), no *creA* mutations were found that have a phenotype of failure to derepress (Arst and Bailey, 1977; Kelly and Hynes, 1977). All

creA mutants obtained to date have a partially derepressed phenotype. For example, the alcohol dehydrogenase and proline oxidase enzyme activities were much higher in a *creA1* strain in the presence of 1% glucose, inducer and ammonium than in wild type (Arst and Bailey, 1977). A convenient assay for the presence of the alcohol dehydrogenase enzyme is provided by the inclusion of allyl alcohol in the growth medium. Allyl alcohol is converted to the toxic compound acrolein by alcohol dehydrogenase (encoded by *alcA*) and a wild type strain is able to grow on glucose in the presence of allyl alcohol because the *alcA* gene is subject to carbon catabolite repression (Pateman *et al.*, 1983; Hynes and Kelly, 1977). Strains containing mutant alleles of *creA* show reduced growth on glucose in the presence of allyl alcohol due to derepression of *alcA* transcription. *creA* mutants display non-heirachical heterogeneity (Arst and Cove, 1973; Bailey and Arst, 1975; Arst and Bailey, 1977). For example, in an *areA1* containing strain *creA2* was able to allow growth on acetamide but not proline whereas *creA25* was able to allow good growth on proline but not on acetamide (Arst and Bailey, 1977). This implies a direct role in carbon catabolite repression for the *creA* gene product, possibly as a DNA binding regulator of gene expression. The *creA* locus was mapped to linkage group I (Bailey and Arst, 1975) and is located between *r*D and *gal*D (Clutterbuck, 1993). *creA* mutations result in affected colony morphology. Some are extremely compact, whereas others are affected less.

In addition to *creA* mutations, Hynes and Kelly (1977) also isolated mutations in a number of other loci. Pleiotropic mutants mapping to genes designated *creB* and *creC* (linkage group II) were identified in the same screen that identified *creA* alleles. These also have a derepressed phenotype for some genes, for example, *amdS*, *alcA, facA* (acetyl CoA synthase) and *acuD* (isocitrate lyase) (Kelly and Hynes, 1977; Hynes and Kelly, 1977). The *creB* and *creC* mutants show similar phenotypes with respect to their utilisation of carbon sources (Kelly, 1980). It has previously been suggested that the *creB* and *creC* gene products have an indirect role in carbon catabolite repression as cell membrane proteins (Arst, 1980). However, given the situation in yeast (section 1.2.2) it is possible that CREB and CREC may be able to interact with CREA, may modify CREA function or they may be involved in signalling the presence of glucose.

The *creD34* mutation (previously called *cre*-34) was isolated as a spontaneous sector of a *creC*27 containing strain able to grow on glucose plus fluoroacetamide and is closely linked (3-6 map units) to *creC* (Kelly and Hynes, 1977). This mutation was able to suppress the *creC27*, *creB15*, and *creA204* phenotypes on glucose plus allyl alcohol media (Kelly and Hynes, 1977; Kelly, 1980). The *creD34* mutation leads to stronger repression of some genes, for example acetyl CoA synthase, compared to wild type, in the presence of a source of carbon catabolite repression (Kelly and Hynes, 1977). All *creA, creB, creC* and *creD* alleles islolated to date are recessive to wildtype alleles.

1.2.3.3 The cloning and preliminary characterisation of the *creA* gene from *A. nidulans*

To investigate carbon catabolite repression at the molecular level in *A. nidulans*, the *creA* gene was cloned by complementation and plasmid rescue. Expression of the *creA* transcript is higher in arabinose grown cultures (derepressing conditions) than in glucose grown cultures (repressing conditions). Transformants containing multiple copies of the *creA* gene grew better than the wild type strain on sucrose medium plus allyl alcohol, indicating tighter repression of the alcohol dehydrogenase gene (Dowzer and Kelly, 1989). The *creA* mutants have higher levels of *creA* mRNA in D-glucose grown mycelia (Dowzer, 1991), which suggests autoregulation may play a role in the control of *creA* expression.

The sequence of the genomic clone and two cDNA clones was determined (Dowzer and Kelly, 1991). The putative protein sequence of 416 amino acids contains several features which are characteristic of regulatory and DNA binding proteins. These include two zinc fingers of the Cys_2His_2 class, an alanine rich region, an acidic region and frequent S(T)PXX motifs (Dowzer and Kelly, 1991). An alignment of the putative CREA protein sequence with sequences in the data base showed a high degree of sequence similarity over the zinc finger region with MIG1 from *S. cerevisiae*, mammalian WT1 protein and the fingers of regulatory protein AMDA from *A. nidulans* (Fig. 1.3, Dowzer and Kelly, 1991).

The alanine rich region, beginning at amino acid position 131 contains nine consecutive alanines (Dowzer and Kelly, 1991). Similar regions are also found in the repressor proteins

Kruppel, Engrailed and Evenskipped from *D. melanogaster*. The alanine rich region in Kruppel has been deleted and this protein was found to have a reduced ability to act as a repressor (Licht *et al.*, 1990). Acidic regions have been associated with transcriptional activation. It is possible that CREA may act as an activator as well as a repressor as is the case for some *Drosophila* proteins, for example Ultrabithorax and the yeast protein MIG1 (Krasnow *et al.*, 1989; Treitel and Carlson, 1995). The *creA* gene from *A. niger* has been cloned and sequenced and an alignment of the two sequences shows a high degree of similarity at the amino acid level (Drysdale *et al.*, 1993). A stretch of 45 amino acids are identical in both species and a search of the data base with this sequence found 33% identity (81% similarity) to a region of the RGR1 protein from *S. cerevisiae* (Drysdale *et al.*, 1993; Sakai *et al.*, 1990). As mentioned in section 1.2.2.3e a strain containing the *rgr1* mutation has a phenotype not dissimilar to a strain containing the *creA* mutation, and so this area of sequence similarity is of potential interest. For a more detailed comparison between MIG1, RGR1 and CREA see section 6.6.

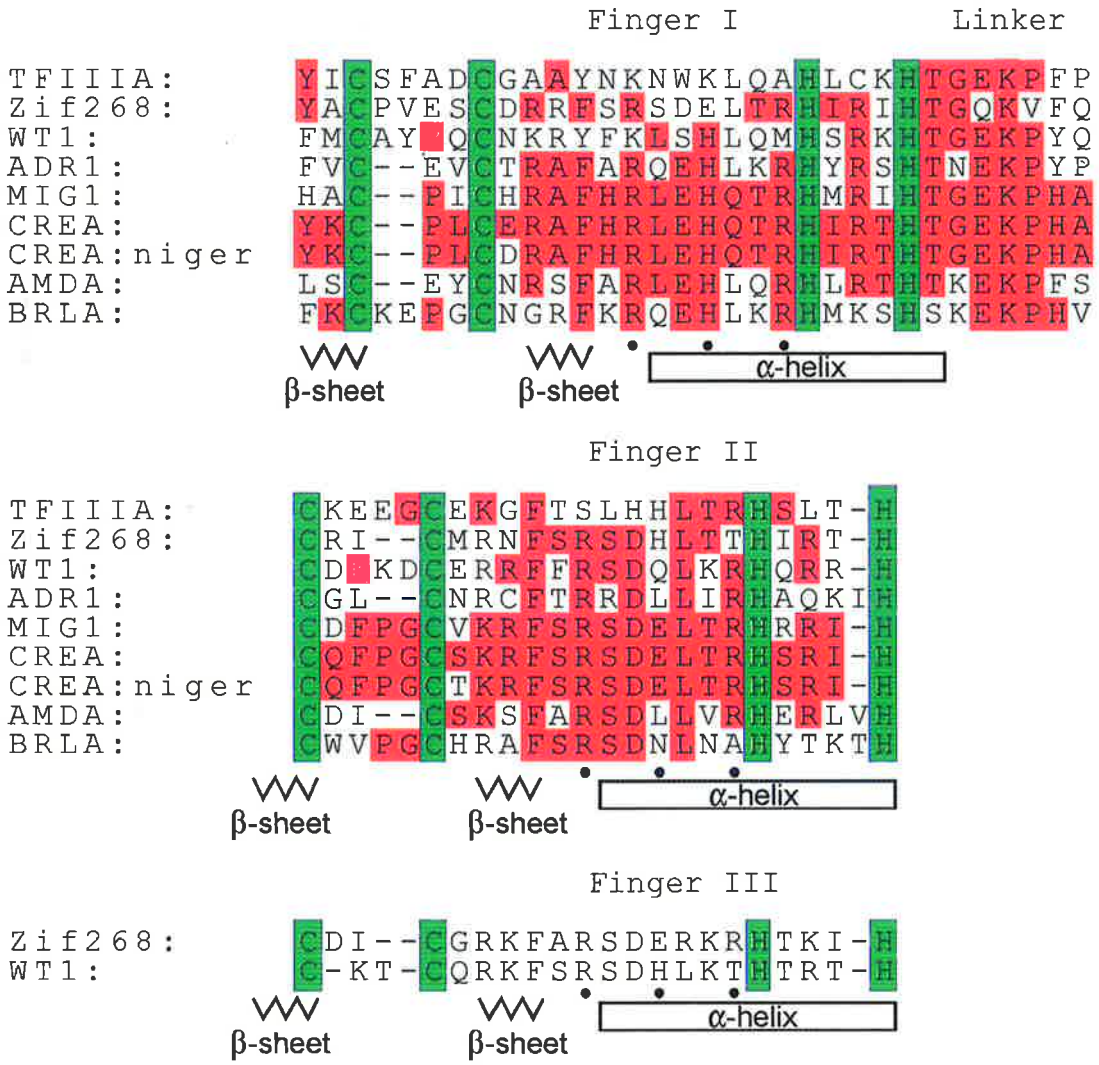
1.2.3.4 The isolation of a strain with a null mutation in *creA*

None of the mutants isolated thus far were shown to be true null alleles. Even *creA30*, a mutation resulting from a pericentric inversion, was shown to contain two mRNA species that hybridise to *creA* (Arst *et al.*, 1990). Dowzer and Kelly (1991) have shown that a deletion of the *creA* locus and adjoining sequences using the cloned gene and *in vitro* techniques, results in a severe reduction in viability (limited spore germination only) unless complemented by another copy of the *creA* gene either in a heterokaryon or a diploid. This was evidence that either total derepression of all functions under *creA* control is lethal or that *cre*A has some essential positive role in gene regulation in *A. nidulans*.

Recently a mutation *creA303* was isolated (M. Hynes, unpublished) which contains a stop codon within the first zinc finger and thus expresses only 69 amino acids of CREA protein (R.A. Shroff and J.M. Kelly, pers. comm.). This strain does not show the severe effect on survival of the deletion of Dowzer and Kelly (1991) which also deleted sequences 5' and 3' of *creA*. It is a formal possibility that it is the deletion of these sequences, rather than *creA*, that is the cause

Figure 1.3 Comparison of the zinc finger motifs from CREA and other regulatory proteins

Comparison of the Cys_2His_2 zinc finger motifs of CREA from *A. nidulans* (Dowzer and Kelly, 1991), CREA from *A. niger* (Drysdale *et al.*, 1993), TFIIIA (Miller *et al.*, 1985), Zif268 (Christy *et al.*, 1988), WT1 (Call *et al.*, 1990), ADR1 (Hartshorne *et al.*, 1986), MIG1 (Nehlin and Ronne, 1990), AMDA (Lints *et al.*, 1995) and BRLA (Adams *et al.*, 1988). The cysteine and histidine residues that co-ordinate the Zn^{2+} cation are boxed in green and the linker region between zinc fingers I and II is marked. The position of β-sheets, the α-helix and amino acids shown to interact with DNA bases in the crystal structure of Zif268 (black dots) is shown (Pavletich and Pabo, 1991). Only the first two finger motifs from TFIIIA and the first three from WT1 are shown. Amino acids identical to those from CREA of *A. nidulans* are boxed in red.


Finger I
Linker
TFIIIA:
Zif268:
WT1:
ADR1:
MIG1:
CREA:
CREA:niger
AMDA:
BRLA:
β-sheet
β-sheet
α-helix
Finger II
TFIIIA:
Zif268:
WT1:
ADR1:
MIG1:
CREA:
CREA:niger
AMDA:
BRLA:
β-sheet
β-sheet
α-helix
Finger III
Zif268:
WT1:
β-sheet
β-sheet
α-helix

of the extreme phenotype. The creation of a more precise deletion has been undertaken and shows a phenotype similar to *creA303* (S.M. O'Connor and J.M. Kelly pers. comm.). The phenotype of a *creA303* mutation is partial derepression of all systems analysed.

1.3 Gene Systems Regulated by *creA* in *A. nidulans*

1.3.1 Regulation of the *amdS* gene

The *amdS* gene of *A. nidulans* encodes an acetamidase which enables the utilisation of acetamide as both a carbon and a nitrogen source by catalysing the hydrolysis of acetamide to acetate and ammonium (Hynes and Pateman, 1970). Expression is induced by acetate and other sources of acetyl-CoA rather than the substrates acetamide (Hynes, 1977) or acrylamide (Hynes and Pateman, 1970). Mutational analysis has shown that the *amdS* gene is subjected to multiple regulatory pathways including global regulatory control by nitrogen metabolite repression (mediated by *areA*) and carbon catabolite repression (mediated by *creA*) as well as many independent induction pathways (reviewed by Davis *et al.*, 1993). Mutations at unlinked loci affecting *amdS* expression have lead to the discovery of *trans* acting factors AMDR, FACB and AMDA which modulate the expression of *amdS* according to the environment. *Cis* acting regulatory mutations define the site of action of *trans* acting factors involved in the induction, activation or repression of the *amdS* gene (Fig. 1.4; Corrick *et al.*, 1987; Hynes *et al.*, 1988).

1.3.1.1 Regulation of *amdS* by *facB*

The *facB* gene encodes a positively acting regulatory protein required for growth on acetate as a sole carbon source. The recessive loss of function mutations *facB101* and *facB501* result in reduced expression of *amdS*, *facA* (acetyl CoA synthase), *acuD* (isocitrate lyase) and *acuE* (malate synthase) (Hynes, 1977, Kelly, 1980). The *acuD* and *acuE* genes encode enzymes of the glyoxylate by-pass. The *cis*-dominant mutation *amdI9* results in greatly increased *facB* dependent induction of the *amdS* gene (Hynes, 1975a; 1977). Titration of the *facB* gene product by multiple copies of the *amdI9* sequence but not the wild type *amdS* sequence is consistent with *facB* binding the *amdI9* sequence with higher affinity (Kelly and Hynes, 1987). An oligonucleotide spanning the *amdI9* mutation is able to confer acetate induction in a strain

Figure 1.4 Regulation of the *amdS* gene from *A. nidulans*

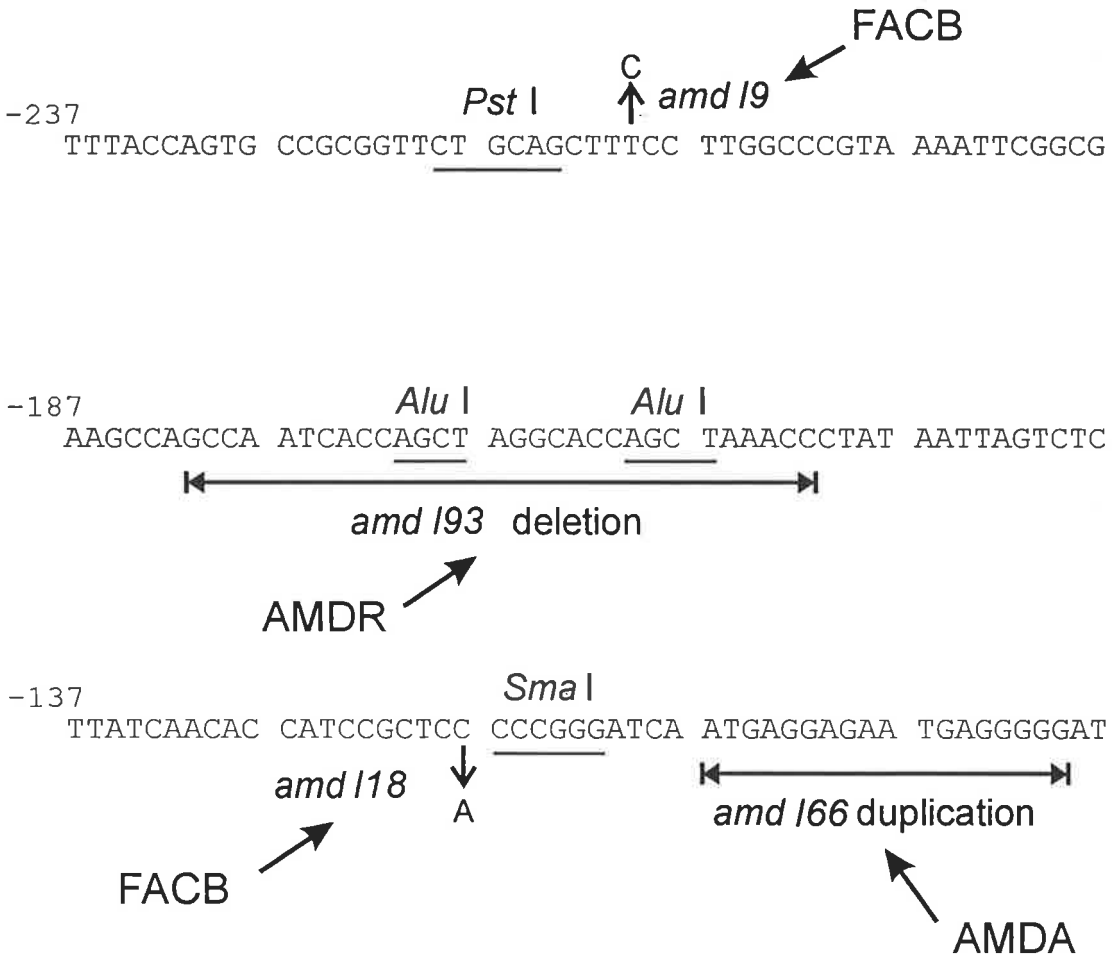
5' sequence of the *amdS* gene showing some restriction endonuclease sites, the position and nature of *cis*-acting mutations and the site of action of some regulatory proteins (Corrick *et al.*, 1987; Hynes *et al.*, 1988). Shown are the following mutations (for a discussion see text):

amdI9: point mutation (T-> C transition at position -210)

amdI93: 30 bp deletion from -152 to -181

amdI18: point mutation (C->A transversion at position -118)

amdI66: 18 bp tandem duplication of the sequence -90 to -181


FACB
C
amd I9
Pst I
-237
TTTACCAGTG CCGCGGTTCT GCAGCTTTCC TTGGCCCGTA AAATTCGGCG
Alu I
Alu I
-187
AAGCCAGCCA ATCACCAGCT AGGCACCAGC TAAACCCTAT AATTAGTCTC
amd I93 deletion
AMDR
Sma I
-137
TTATCAACAC CATCCGCTCC CCCGGGATCA ATGAGGAGAA TGAGGGGGAT
amd I18
A
amd I66 duplication
FACB
AMDA
-87
GCGGGGCTAA AGAAGCCTAC ATAACCCTCA TGCCAACTCC CAGTTTACAC
Alu I
-37
+1
TCGTCGAGCC AACATCCTGA CTATAAGCTA ACACAGAATG CCTCAA
Met Pro Gln

transformed with a construct containing 5' *amdS* (deleted) fused to the reporter gene β-galactosidase, where the wild type copy of *amdS* was replaced in the two step transformation procedure of Davis *et al.*, (1988) (M.J. Hynes pers. comm.). Mutations in *facB* have recently been found to be epistatic to the up-promoter mutation *amdI18* implying a role for this region in *facB*-dependent activation of *amdS* (M.J. Hynes pers. comm.; Hynes, 1978a). The *facB* gene has been cloned (Katz and Hynes, 1989b) and sequenced (Todd *et al.*, submitted) and encodes a $Zn(II)_2Cys_6$ zinc cluster DNA binding motif (section 1.1.2.3), a putative leucine zipper dimerisation domain, and two acidic α-helical regions which may be involved in activation (Todd *et al.*, submitted; Todd, 1995). The FACB protein has been shown to bind to three regions 5' of *amdS*, including the regions surrounding the *amdI9* and *amdI18* mutations and to regions 5' of the *facA*, *facB*, *acuD* and *acuE* genes (Todd, 1995). FACB binding sites were found to fall into two classes with the following consensus sequences: 5' T C C/G N_{8-10} C/G G A 3' (class A) and 5' G C A/C N_{8-10} G/T G C 3' (class B) (Todd, 1995).

The semi-dominant gain of function *facB88* mutation results in high levels of constitutive expression of *amdS* but normal expression of the genes of acetate utilisation (Katz and Hynes, 1989b, M.J. Hynes unpublished). The *facB88* mutation is a reciprocal translocation between *facB* and a gene called *amdX*. Both breakpoints have been cloned and a hybrid gene containing the 5' half of *facB* fused to *amdX* is sufficient to confer super activation of *amdS* (M.J. Hynes pers. comm.). The *amdX* gene has been sequenced and encodes two Cys_2His_2 zinc finger DNA binding motifs (Murphy, 1996). Binding studies using a glutathione-S-transferase (GST)-AmdX fusion protein have shown that AMDX binds strongly to the regions -104 to -74 and -109 to -87 5' of *amdS*. The predicted 5' FACB-AMDX protein encoded by the *facB88* containing strain contains both the FACB DNA binding motif and the AMDX zinc finger DNA binding motif (Murphy, 1996). Using various 5' *amdS-LacZ* fusion constructs the *facB88* translocation was shown to activate transcription when either the FACB or the AMDX binding sites were present, however, both types of binding sites were required for super activation of the reporter gene construct (Murphy, 1996). These results will be discussed further in section 6.6.3.3.

1.3.1.2 Regulation of *amdS* by *amdA*

A semi-dominant mutation within the *amdA* gene (*amdA7)* was isolated, as an acetamide specific suppressor of the *areA217* allele, that caused elevated levels of *amdS* expression (Hynes, 1975b, 1978b). This effect was greatly increased by the *cis*-acting mutation *amdI66* (Hynes, 1982). More recently a *cis*-acting mutation *amdI666* was isolated which results in increased levels of acetamidase even in a wild type *amdA* background (Katz *et al.*, 1990). The *amdI66* and *amdI666* mutations are duplications and triplications respectively of a GA rich sequence 5' of *amdS* (Hynes *et al.*, 1988; Katz *et al.*, 1990). Similar GA rich sequences are present in the 5' region of the *aciA* gene, a putative formate dehydrogenase which is subject to *amdA* mediated induction by acetate (Saleeba *et al.*, 1992; Atkinson *et al.*, 1985; Chow and RajBhandary, 1993). The *amdA7* mutation was also shown to affect the expression of *alcB* (Sealy-Lewis, 1990) although a metabolic link between ALCB, ACIA and AMDS is unknown. The *amdA* gene product is only a minor contributor to *amdS* induction by acetate since an *amdA* loss of function mutation does not have a large effect on the expression of *amdS* (Hynes, 1982; Atkinson *et al.*, 1985).

The *amdA* gene has been cloned and the deduced amino acid sequence contains two putative zinc fingers of the Cys_2His_2 type with 53 % sequence similarity at the amino acid level with CREA (Fig. 1.3, Lints *et al.*, 1995). The putative protein contains two proline rich regions and an acidic region, both of which may function as activation domains (Lints *et al.*, 1995). Deletion studies of 5' *amdS-LacZ* constructs has located the binding of AMDA to between -75 and -101 5' of the start point of translation of *amdS* (Lints *et al.*, 1995). This coincides with the region duplicated and triplicated in the *amdI66* and *amdI666* mutations respectively. A strain containing a large deletion of *amdA* sequence is viable proving *amdA* is not required for any essential functions (Lints *et al.*, 1995). The *amdA* gene was shown to be constitutively expressed under conditions that affect expression of *amdS* and *aciA* (Lints *et al.*, 1995).

1.3.1.3 The regulation of *amdS* by *amdR*

The acetamidase is induced by ω-amino acids such as β-alanine and γ-amino butyric acid (GABA), even though the metabolism of these compounds does not require the *amdS* gene

product (Arst, 1976). The *amdR* gene encodes a positively acting regulatory protein which mediates the expression of *amdS* and the genes involved in ω-amino acid metabolism namely GABA transaminase (*gatA)* , GABA permease (*gabA*) and lactamase and/or lactam permease (*lamA* and *lamB*) in response to the presence of ω-amino acids (Arst, 1976; Arst *et al.*, 1978; Katz and Hynes, 1989a; Richardson *et al.*, 1989). Within the 5' region of *amdS* the *cis*-acting mutation *amdI93* abolishes induction by ω-amino acids but not other inducers and is epistatic to constitutive *amdR* alleles (Hynes, 1980).

The AMDR protein contains a DNA-binding domain of the Cys_6 zinc binuclear cluster class (Andrianopoulos and Hynes, 1990) and AMDR binds *in vitro* to the 30 bp region deleted in the *amdI93* mutation. The *A. nidulans* "CCAAT"-binding factor (AnCF) which does not bind a mutated CCTTT box, was also found to bind this region (van Heeswijck and Hynes, 1991). In addition the AMDR protein has been shown to bind to the 5' region of the *gatA*, *lamA* and *lamB* genes with the consensus 5' T T C G G C G N_7 C C A A T 3' (Richardson *et al.*, 1992). Titration of the AMDR protein with multiple copies of the AMDR binding site was shown and results in poor growth on substrates like 2-pyrrolidinone which requires the expression of these genes for metabolism (Kelly and Hynes, 1987; Richardson *et al.*, 1989, 1992).

1.3.1.4 Regulation of *amdS* by *areA*

amdS is repressed by nitrogen metabolites such as ammonium, L-glutamate and L-glutamine (Hynes, 1974) and this is mediated by the positive regulator encoded by the *areA* gene (discussed in section 1.2.3.1). The *areA* gene encodes a Cys_4 zinc finger protein with significant sequence similarity to NIT2 from *N. crassa* (section 1.1.2.2; Fu and Marzluf, 1990a; Kudla *et al.*,1990). The *nit-2* gene can complement an *areA* loss of function mutation and thus AREA is likely to recognise very similar sequences (Davis and Hynes, 1987). Footprinting studies have shown that NIT2 recognises the sequence 5' TATCTA 3' within NIT3 (nitrate reductase) from *N. crassa*, which contains the GATA sequence on the non-coding strand, and also the four TATCTA sequences present in the *A. nidulans niiA-niaD* intergenic region (Fu and Marzluf,

1990b). Using cyclic amplification and selection of targets (CASTing) to select oligonucleotides that contain binding sequences for AREA or the mammalian GATA-1,2,3 or 4 proteins Merika and Orkin (1993) found that AREA also binds the GATA sequence. These five proteins were able to recognise sequences that differed from the WGATAR consensus and they each exhibited distinct preferences for recognition sites. The single fingered AREA protein recognised sequences most similar to the two fingered GATA-1 vertebrate protein. Peters and Caddick (1994) have found that native AREA binds to DNA fragments containing two GATA sequences with higher affinity than DNA fragments with just one recognition site. This effect was independent of the orientation of the two sites. There were no *cis* acting mutations 5' of *amdS* which affected *areA* regulation of *amdS*. For this reason a functional analysis has centred on *in vitro* mutagenesis of the potential GATAA like site 5' of *amdS*. Mutation of this site drastically reduced *amdS* expression specifically under nitrogen limiting conditions but did not totally eliminate it. This suggests that there may be other functional sites in the *amdS* 5' region (M.J. Hynes and M.A. Davis unpublished, cited by Davis *et al.*, 1993).

1.3.1.5 Regulation of *amdS* by *creA*

The *amdS* gene is repressed by glucose and other sources of carbon catabolite repression (Hynes, 1970) and this is mediated by the wide domain regulatory gene *creA*. Mutant strains such as *creA204* are partially derepressed for acetamidase production in the presence of glucose (Hynes and Kelly, 1977; Kelly and Hynes, 1977). Repression of *amdS* in glucose growth conditions was shown to be at the level of transcription (Hynes *et al.*, 1983) and it is the aim of this study to investigate the molecular mechanism behind this repression.

No *cis* acting mutations resulting in either derepression or failure to derepress were found despite the use of strong selection procedures. Therefore it is possible that CREA does not have a direct role in carbon catabolite repression of *amdS*, but rather that repression of the positively acting regulatory gene *facB* explains the repression of *amdS*. The limitation of either carbon or

nitrogen is sufficient to allow *amdS* expression and therefore it is of great interest to determine the molecular mechanism behind the interaction between these two global regulatory pathways.

1.3.1.6 Other regulators of *amdS*

There is evidence that benzamide and benzoate are inducers of the *amdS* gene even though metabolism of these two compounds is not known to require the *amdS* gene product (Hynes, 1978b). Induction by these two compounds is not known to proceed via the same pathway as any of the *trans* acting factors discovered so far.

The CCAAT sequence 5' of *amdS* is known to be important in setting both basal and derepressed levels of *amdS* expression. As discussed in section 1.2.2.2b the HAP2/3/4/5 protein complex is able to bind CCAAT sequences. In order to determine whether the *S. cerevisiae* HAP2/3/4/5 protein complex is able to recognise the CCAAT sequence present in the 5' region of *amdS* various 5' *amdS* fragments were cloned into the *CYC1-lacZ* reporter construct and transformed into wild type and *hap2* yeast strains (Bonnefoy *et al.*, 1995). The results show that *amdS* sequences are capable of regulated expression in yeast in response to carbon limitation and that the HAP2/3/4/5 complex is able to recognise the *amdS* CCAAT sequence (Bonnefoy *et al.*, 1995). These results suggest that AnCF may be functionally related to the HAP system in yeast.

1.3.2 The regulation of the *facB* gene

Northern experiments show that the *facB* transcript is acetate inducible and carbon catabolite repressible (Katz and Hynes, 1989b). Furthermore, carbon catabolite repression was alleviated in a strain carrying the *creA204* mutation and this was seen at the level of mRNA accumulation. The induction of *facB* by acetate was much reduced in a *facB101* loss of function mutant, suggesting an autoregulatory role for FACB (Katz and Hynes, 1989b). Binding by FACB to the 5' *facB* region supports the role of FACB in positive autoregulation (Todd, 1995). The presence of temperature sensitive *facB* mutations which result in the thermolability of enzymes of the

glyoxylate by-pass, suggests that FACB has a structural as well as a regulatory role (Kelly, 1980).

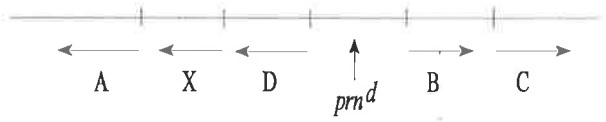
1.3.3 The regulation of the ethanol utilisation regulon of *A. nidulans*

The ability to utilise ethanol as a sole carbon source requires the *alcA* (alcohol dehydrogenase), *aldA* (aldehyde dehydrogenase) and *alcR* (positive regulatory gene) genes. *A. nidulans* has a large number of additional alcohol dehydrogenase genes, including *alcB* (unknown physiological function) and *alcC* (involved in long term survival of anaerobic stress) (Sealy-Lewis and Lockington, 1984; Kelly *et al.*, 1990; reviewed by Felenbok and Sealy-Lewis, 1994). ALCR is required for the induction of the structural genes *alcA* and *aldA* and Northern analysis has shown that this regulation is at the level of mRNA accumulation (Lockington *et al.*, 1985; reviewed by Felenbok and Sealy-Lewis, 1994). As discussed previously, expression of the *alcA* gene is subject to carbon catabolite repression since mutations at the *creA* locus result in toxicity to allyl alcohol in the presence of glucose (Hynes and Kelly, 1977). The *creA1* mutation results in increased *alcR, alcA* and *aldA* mRNA expression in the presence of glucose and an inducer, and this is consistent with a direct role for CREA in the regulation of this regulon (Lockington *et al.*, 1987). It was possible that carbon catabolite repression of the structural genes is mediated by repression of the positively acting regulatory gene *alcR*, at the level of transcription. However, when *alcR* is constitutively expressed under the *gpdA* promoter, some repression of the *alcA* and complete repression of the *aldA* mRNA was seen (Mathieu and Felenbok, 1994; Kulmburg, 1991, cited by Felenbok and Sealy-Lewis, 1994; discussed further in section 6.6.3.1). This suggests that a double block mechanism of repression acts on the regulatory gene *alcR* and the structural genes *alcA* and *aldA*. The ALCR protein contains a Cys_6 binuclear cluster DNA binding motif of the GAL4 type (section 1.1.2.3; Kulmburg *et al.*, 1991). Mobility shift and footprinting analysis has shown that a GST-ALCR fusion protein binds to target sites in the 5' region of *alcR* and *alcA* but not *aldA* (Kulmburg *et al.*, 1992a,b; Kulmburg, 1991, cited by Felenbok and Sealy-Lewis, 1994). The lack of ALCR binding to the 5' *aldA* region may indicate that *aldA* is not directly regulated by *alcR*. The binding of CREA to the 5'

regions of *alcR* and *alcA* (Kulmburg *et al.*, 1993) will be discussed in section 3.9.5.1.

1.3.4 The regulation of the proline utilisation cluster of *A. nidulans*

The proline (*prn*) utilisation cluster of *A. nidulans*, on chromosome VII, contains an internal regulatory region (Arst and MacDonald, 1975) between divergently transcribed genes (V. Gavrias and C. Scazzocchio unpublished data, cited by Sophianopoulou *et al.*, 1993, see below). Expression of the proline oxidase (*prnD*), L-Δ^1-pyrroline-5-carboxylate dehydrogenase (*prnC*) and to a lesser extent the proline transport system (*prnB*), is dependent on the presence of a functional positive regulator PRNA which mediates induction by proline (Sharma and Arst, 1985; Sophianopoulou *et al.*, 1993). *prnX* is a newly identified gene of unknown function.



The proline utilisation cluster is also subject to control by nitrogen metabolite and carbon catabolite repression (Arst and MacDonald, 1975). A strain carrying an *areA*$^-$ mutation is unable to grow on proline as a nitrogen source when glucose is present due to carbon catabolite repression of the genes in the proline utilisation cluster. Early mutational work isolated two classes of mutants able to utilise proline as a nitrogen source in the presence of a repressing carbon source. One class of mutation was specific for proline utilisation while the other mutations affected carbon catabolite repression, mapping to the *creA* locus (Arst and Cove, 1973). The three proline specific mutations *prn*^{d}*-20, prn*^{d}*-21* and *prn*^{d}*-22* map to the internal regulatory region of the *prn* cluster, located between *prnD* and *prnB* (Arst and MacDonald, 1975; Arst *et al.*, 1977; Sophianopoulou *et al.*, 1993). Classical *cis-trans* tests show that the *prn*^{d} mutations result in derepression of the proline transport system (*prnB*) directly (Arst and MacDonald, 1975). Whether they affect the expression of the other genes in the cluster directly is not known since it is possible that *prnB* expression is the limiting factor for proline utilisation (Arst *et al.*, 1980). It was shown that the *prn*^{d}*-20* and *prn*^{d}*-22* mutant strains had higher levels

of *prnB, prnC* and *prnD* mRNAs, compared to the wild type, when grown in the presence of glucose and the inducer proline (Sophianopoulou *et al.*, 1993). Unlike the situation for the ethanol regulon, the *prnA* gene is not subject to carbon catabolite repression (Sophianopoulou *et al.*, 1993). Sequencing of the prn^d mutations and binding studies with the CREA protein confirm that these mutations define the site of action of CREA and these results will be discussed further in section 3.9.5.2.

1.4 Aims and Objectives

The preceding sections described the importance of *creA* in carbon catabolite repression and a study of its function is vital to further our understanding of the molecular means by which carbon catabolite control is exerted. Before this study commenced in 1990 the *creA* gene from *A. nidulans* had been cloned by complementation of the mutant allele *creA204* (Dowzer and Kelly, 1989). Sequencing of a genomic clone and two cDNA clones was underway. Analysis of the putative *creA* sequence revealed two zinc finger motifs of the Cys_2His_2 type (Dowzer and Kelly, 1991). The presence of this motif suggested a possible role for the *creA* gene product as a DNA binding protein. A direct role for CREA in controlling the expression of carbon catabolite repressed genes at the level of transcription is investigated here by an analysis of binding by CREA to the 5' regions of genes under CREA control. The aim of this study was to investigate the function of CREA in the molecular mechanism of carbon catabolite repression, with particular reference to the *amdS*, *facB* and *creA* genes of *A. nidulans*.

CHAPTER 2

Chapter 2

Materials and Methods

2.1 Materials

General reagents: General chemicals, media requirements and supplements were of laboratory grade and were purchased from various companies.

Nucleotides and isotopes: α-^{32}P-dATP (3000 Ci/ mmole) and γ-^{32}P-dATP (4000Ci/ mmole) were purchased from Bresatec Pty. Ltd. $^{35}SO_4^{2-}$ was purchased from Amersham.

Molecular weight markers: DNA molecular weight standards (λ/*Hind* III, SPP-1 bacteriophage/*Eco* RI and pUC19/*Hpa* II) were purchased from Bresatec Pty. Ltd. Protein molecular weight markers (pre-stained) were purchased from Sigma.

Enzymes: Enzymes for general DNA manipulations were purchased from Boehringer-Mannheim GmbH, Progen Industries or New England Biolabs. *Taq* polymerase and the Klenow fragment of *E. coli* DNA polymerase I was purchased from Bresatec and Novozyme was purchased from Novo Industries.

2.1.1 Buffers

Restriction endonuclease buffers: Digest buffers or their details were supplied by the manufacturer(s)

10 X Ligation buffer (for 5' and 3' overhangs): 100 mM $MgCl_2$, 100 mM DTT, 10 mM spermidine, 10 mM ATP, 500 mM Tris.HCl (pH 7.5)

10 X Blunt end ligation buffer: 200 mM Tris.HCl (pH 7.6), 50 mM $MgCl_2$, 50 mM DTT

10 X Polynucleotide kinase buffer: 500 mM Tris.HCl (pH 7.5), 100 mM $MgCl_2$, 50 mM DTT

10 X Nick translation buffer: 500 mM Tris.HCl (pH 7.5), 75 mM MgAc, 40 mM DTT, 1 mg/ ml BSA

10 X Stop buffer for nick translations: 10 mM Tris.HCl (pH 7.5), 1 mM EDTA, 2% SDS

1 X Loading buffer for agarose gels: 0.25% bromophenol blue, 0.25% xylene cyanol FF, 25% ficoll (type 400)

DNase I dilution buffer: 50 mM Tris.HCl, 5 mM $CaCl_2$, 20% glycerol, pH 6.5

Immunoprecipitation buffer: 1% triton X-100, 0.137 M NaCl, 1 mM $CaCl_2$, 10% glycerol, 20 mM Tris.HCl, pH 9.0

2.1.2 Solutions

2 X M9 Salts: (per litre) 12 g Na_2HPO_4, 6 g KH_2PO_4, 1 g NaCl, 2 g NH_4Cl

1 X SSC: 0.15 M NaCl, 0.015 M $Na_3C_6H_5O_7.2H_2O$ (trisodium citrate), pH 7.2

1 X SSPE: 0.18 M NaCl, 10 mM NaH_2PO_4, 1 mM EDTA, pH 7.4

1 X STE: 10 mM Tris.HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA

1 X TAE: 40 mM Tris base, 20 mM NaAc, 2 mM EDTA, pH 7.8 with glacial acetic acid

1 X TE: 10 mM Tris.HCl (pH 8.0), 1 mM EDTA

PBS: 150 mM NaCl, 16 mM Na_2HPO_4, 4 mM NaH_2PO_4, pH 7.3

10 X TBE (Method I): 1 M Tris base, 0.9 M boric acid (H_3BO_4), 0.2 M EDTA, pH should be 8.3 when diluted

10 X TBE (Method II): 0.9 M Tris base, 0.9 M boric acid, 0.25 M EDTA, pH should be 8.3 when diluted

Aspergillus trace element solution: (per litre) 40 mg $Na_2B_4O_7$, 400 mg $CuSO_4$, 1 g $FePO_4$, 600 mg $MnSO_4.H_2O$, 800 mg $Na_2MoO_4.2H_2O$, 8 g $ZnSO_4.7H_2O$ and 2 ml $CHCl_3$ as preservative.

Aspergillus vitamin solution: (per litre) 40 mg *p*-aminobenzoic acid, 50 mg thiamine.HCl, 1 mg D-Biotin, 400 mg inositol, 100 mg nicotinic acid, 200 mg calcium D-pantothenate, 100 mg riboflavin, 50 mg pyridoxine and 2 ml $CHCl_3$ as preservative.

Aspergillus salt solution: (per litre) 26 g KCl, 26 g $MgSO_4.7H_2O$, 76 g KH_2PO_4, 50 ml *Aspergillus* trace element solution and 2 ml $CHCl_3$ as preservative. pH should be 4.0.

2.1.3 Media

Bacterial growth media:

L-broth: 1% NaCl, 0.5% yeast extract, 1% tryptone, pH 7.2. Plates were solidified with 1.0% agarose or Oxoid class I agar.

M9 glucose minimal medium: 50% 2 X M9 salts, 10 mM $MgSO_4$, 1 mM $CaCl_2$, 10 mM thiamine.HCl, 0.2% D-glucose. Plates were solidified with 1.5% Oxoid Class I agar.

SOC: L-broth plus 10 mM $MgSO_4.7H_2O$, 10 mM $MgCl_2$, 2.5 mM KCl, 1% glucose.

Antibiotics: Antibiotics were added to media to give a final concentration of:

ampicillin	50 μg/ ml
chloramphenicol	0.17 mg/ ml

Aspergillus growth media:

Sucrose medium: 2% *Aspergillus* salt solution, 1% sucrose, pH 6.5. Unless otherwise indicated, nitrogen sources were added to a final concentration of 10 mM. Plates were solidified with 1% agar.

Carbon free medium: 2% *Aspergillus* salt solution, pH 6.5. Plates were solidified with 1 or 2.2 % agar. Unless otherwise indicated, carbon sources were added to a final concentration of 1% and nitrogen sources added to a final concentration of 10 mM.

Complete medium: 1% D-glucose, 0.2% peptone, 0.15% casein hydrolysate, 0.1% yeast extract, 10 mM ammonium (+)-tartrate, 2% *Aspergillus* salt solution, 1% *Aspergillus* vitamin solution, 25 μg/ml riboflavin, pH 6.5. Plates were solidified with 1 or 2.2 % Oxoid Class III agar.

Aspergillus nitrogen free medium (glucose medium): 2% *Aspergillus* salt solution, 1% D-glucose, pH 6.5. Unless otherwise indicated, nitrogen sources were added to a final concentration of 10 mM. Plates were solidified with 1 or 2.2 % agar.

Protoplast medium: 1 M sucrose, 1% D-glucose, 2% *Aspergillus* salt solution, pH 7.0 and solidified with 1% or 0.25% agar for underlayer and overlayer of plates respectively.

Unless otherwise indicated, nitrogen sources were added to a final concentration of 10 mM.

Low sulfate minimal medium (per litre): 10 g D-glucose, 1 g K_2HPO_4, 0.5 g $MgCl_2.6H_2O$, 0.5 g KCl, 0.01 g ferrous sulfate, pH 6.8 with HCl.

Supplements: Growth supplements were added as required to give the following concentrations:

L-arginine	0.12 mg/ml
D-biotin	0.01 µg/ml
p-aminobenzoic acid	50 µg/ml
nicotinic acid	1.0 µg/ ml
pyridoxine sulfate	0.5 µg/ ml
riboflavin	2.5 µg/ml

2.1.4 *E. coli* Strains: *E. coli* strain DH1 (*supE*44, *hsd*R17, *recA*1, *endA*1, *gyrA*96, *thi*-1, *relA*1) (Hanahan, 1983) was used for general plasmid maintenance. *E. coli* strain JM101 (genotype: *supE*, *thi*Δ(*lac-proAB*), F'[*traD*36, *proAB*$^+$, *lacI*^{q}, *lacZ*M15])(Messing, 1979) was used to screen recombinant plasmid. *E. coli* protein extracts (section 2.2.2) were made from protease deficient (*lon*$^-$) BL21(DE3) cells (genotype: *hsdS*, *gal* (λc*Its*857, *ind*1, *S*am7, *nin*5, *lac*UV5-T7 gene *1*) (Studier and Moffatt, 1986) transformed with pGEX-2T based constructs.

2.1.5 Fungal strains: A list of *A. nidulans* strains used in this study and their genotypes, is given in Table 2.1. For meanings of gene symbols see Clutterbuck (1993).

2.1.6 Plasmids: The plasmids used in this study are presented in Table 2.2.

2.1.7 Oligonucleotides: Oligonucleotides used in this study are presented in Table 2.3. Unless otherwise stated oligonucleotides were purchased from the Microbiology Department, University of Adelaide or synthesised on a Beckman model T1000 DNA synthesiser according to the manufacturers instructions.

Table 2.1 *A. nidulans* strains used in this study

Strain	Genotype	Reference
"wild type"	*biA1; niiA4*	Cove and Pateman, (1963)
C43	*biA1 pabaA1 creA204; argB2*	Dowzer, (1991)
664	*biA1 creA204; niiA4*	Hynes and Kelly, (1977)
CR1	C43::pANC6 pMOO6	this thesis
CR6	C43::pANC6 pMOO6	this thesis
11	C43::pANC6 pMOO6	this thesis
creA30	*biA1 creA30*	Arst *et al.*, (1990)
creA218	*yA1 creA218; areA217; riboB2*	M.J. Hynes, unpublished, Shroff *et al.*, (submitted)
creA220	*yA1 creA220; areA217; riboB2*	M.J. Hynes, unpublished, Dowzer and Kelly, (1989)
creA221	*yA1 creA221; areA217; riboB2*	M.J. Hynes, unpublished, Shroff *et al.*, (submitted)
creA303	*yA1 creA303; areA217; riboB2*	M.J. Hynes, unpublished
creA304	*yA1 creA304; areA217; riboB2*	M.J.Hynes, unpublished

Table 2.2 Plasmid clones of *A. nidulans* DNA used in this thesis

Plasmid	Insert	Reference
pCD5	cDNA clone of *creA*	Dowzer and Kelly, (1991)
pGEX-*creA*	621 bp *Nco* I-*Nco* I fragment from pCD5 cloned in frame into the *Sma* I site of pGEX-2T	Kulmburg *et al.*, (1993)
pGEX-*creA*(MF)	621 bp end filled *Nco* I-*Nco* I fragment from pMF cloned in frame into the *Sma* I site of pGEX-2T	this thesis
pGEX-*creA*(RGR)	The 474 bp *Eco* RV-*Eco* RV fragment from pANC6 inserted into end filled pGEX-2T digested with *Eco* RI	this thesis
pANC4	2.3kb *Bam* HI-*Xba* I genomic clone of *creA*	Dowzer and Kelly, (1989)
pANC6	7.5kb *Eco* RI genomic clone of *creA*	Dowzer (1991)
pMF	pANC4 equivalent of *creA204* with 3 PCR induced errors (in pUC19)	R.A. Shroff and J.M. Kelly, pers. comm.
p3SR2	5kb *Eco* RI-*Sal* I fragment containing the *amdS* gene	Hynes *et al.*, (1983)
pFAB5-1 Sac SK+-9HIII self-4	*Sac* I-*Hind* III fragment containing 569 bp of 5' region from *facB* in pBLUESCRIPT	M.J. Hynes unpublished
pFAB5-1	4kb *Sca* I-*Sca* I fragment containing 1.4 kb of 5' region from *facB*	Katz and Hynes, (1989b)
pLIT14	*Bam* HI linker inserted into the *Sma* I site of the *amd*I9-*lacZ* construct	Littlejohn and Hynes, (1992)
p718B16-322	5kb *Eco* RI-*Sal* I fragment containing the *amdI18* mutation from *amdS*	Corrick *et al.*, (1987); Hynes, (1978a)
pMOO6	*argB* gene	Upshall, (1986)

Table 2.3 Oligonucleotides used in this study

Name	Sequence
31 /32 *	5'GATCAACACCATCCGCTC CCCCGGG ATCAATG 3' 3' GTTGTGGTAGGCGAG GGGGCCC TAGTTACTAG 5'
RL2/ RL3 *	5' GAT CCA TGA GGA GAA TGA GGG GGA TG 3' 3' GT ACT CCT CTT ACT CCC CCT ACC TAG 5'
ONC1 /ONC2	5' GATC AG GAG AAT GA**G GGG** GAT GC**G GGG** CTA AAG AA 3' 3' TC CTC TTA CTC CCC CTA CGC CCC GAT TTC TT 5'
ONC4	5' GAT CAG GAG AAT GAG **T**GG GAT GCG **T**GG CTA AAG AA 3'
ONC5	5' GAT CAG GAG AAT GAG G**A**G GAT GCG G**A**G CTA AAG AA 3'
ONC6	5' AG AAT GAG GGG GAT GCG **T**GG CTA AAG AA 3'
ONC7	5' TT CTT TAG C 3'
ONC10	5' GAT CAG GAG AAT GAG **T**GG GAT GCG GGG CTA AAG AA 3'
ONC12	5' TCG GAT CCC GTT TTT AGC 3'
ONC13	5' TGG AAT TCT AGG CAC 3'
687	5' AATTGAACGC N_{20} GGTTACACCA 3'
602	5' AACG AATTGAACGC 3'
603	5' TCAA TGGTGTAACC 3'

* kindly supplied by M. J. Hynes (unpublished)

2.2 Methods

General molecular biological methods, for example, Southern and Northern transfers (Using Zeta-Probe®), hybridisations, and agarose and polyacrylamide gel electrophoresis, were carried out as indicated by the manufacturers or as outlined in laboratory manuals (Maniatis *et al.*, 1982, Sambrook *et al.*, 1989, Ausubel *et al.*, 1987). Double stranded sequencing was carried out using a Sequenase^{R} T7 DNA polymerase version 2.0 kit from USB and thermocycle sequencing was carried out using a *f*mol kit from Promega.

2.2.1 Isolation of plasmid DNA

Small scale preparation of plasmid DNA was carried out by the method of Ish-Horowicz and Burke, (1981). Large scale preparation of plasmid DNA was performed by the alkaline lysis method of Sambrook *et al.*, (1989), except that plasmid DNA was usually resuspended in H_2O instead of 1 X TE.

2.2.2 Isolation of *E. coli* total protein extracts

Proteins were isolated from *E. coli* as outlined in Smith and Johnson (1988). *E. coli* BL21 cells transformed with either pGEX-2T, or a recombinant pGEX-2T construct containing the region of interest, were grown to saturation in 25 ml L-broth + ampicillin. This culture was used to seed a 500 ml L-broth containing ampicillin and 0.6 mM IPTG. After 3 hours of growth, cells were harvested by centrifugation at 10,400g for 10 min, resuspended in 5 ml cold PBS and placed on ice before being sonicated using an MSE 150 watt ultrasonic disintegrator for 4 min. 1 ml of 10% triton X-100 was then added and the mixture transferred to a corex tube and spun at 20,000g for 10 min. The supernatant was filtered through a disposable 0.2 μm filter. Protein concentrations were determined by the method of Bradford (1976) using commercially available reagents (Bio-Rad laboratories). Protein extracts were stored as 1 ml aliquots at -80 °C.

2.2.3 Affinity column chromatography of *E. coli* protein extracts

1 ml of glutathione sepharose 4B beads (Pharmacia) was placed inside a 5 ml syringe with Whatman glass microfibre filter as support. All solutions were filtered through a 0.2 μm filter before loading onto the column. The column was washed with 10 ml of PBS and equilibrated with 10 ml of PBS+ 1% triton X-100.

0.5 ml of *E. coli* total protein extract derived from cells harbouring the pGEX-2T plasmid or from cells harbouring recombinant plasmid was diluted in 10 ml of PBS and then applied to the column five times and the final filtrate discarded. The column was rinsed with 10 ml of PBS. The proteins were eluted with 10 ml of 50 mM reduced glutathione (Sigma), 4 mM Tris.HCl. The protein profiles of the fractions were determined by SDS PAGE (section 2.2.4) and total protein assays were carried out by the method of Bradford (1976) using reagents purchased from Bio-Rad laboratories.

2.2.4 SDS PAGE (polyacrylamide gel electrophoresis)

Discontinuous SDS polyacrylamide gels were electrophoresed at 80 v for 2.5 h or until the bromophenol blue tracker dye reached the bottom of the gel.

2 X SDS Loading buffer: 20 mM Tris buffer pH 6.8, 33% glycerol, 3.3% SDS, 50 mM β-mercaptoethanol, 0.02% bromophenol blue

Separator gel: 3.3 ml 30% acrylamide /0.8% N,N'-methylene bisacrylamide, 2.5 ml 4 X Tris/SDS buffer pH 8.8, 4.2 ml dH_2O, 33 mg ammonium persulfate, 20 μl TEMED

Stacker gel: 1.25 ml of 10% acrylamide/ 2.5% N,N'-methylene bisacrylamide, 1.25 ml 4 X Tris.HCl/ SDS buffer pH 6.8, 3.05 ml dH_2O, 25 mg ammonium persulfate, 5 μl TEMED

4 X Tris.HCl/SDS Buffer pH 8.8: 1.5 M Tris base, 0.4% SDS, pH 8.8 with HCl, filter through 0.2 μm filter

4 X Tris.HCl/SDS Buffer pH 6.8: 0.5 M Tris base, 0.4% SDS, pH 6.8 with HCl, filter through 0.2 μm filter

5 X Tris/Glycine/SDS Running Buffer: 0.125 M Tris base, 1 M glycine (amino acetic acid), 0.5% SDS, pH when diluted should be 8.3

Fixing Solution: 10% glacial acetic acid, 10% methanol

Coomassie blue staining solution: 50% methanol, 0.05% Coomassie brilliant blue^{R}, 10% acetic acid. Coomassie blue was dissolved in methanol first.

Destaining solution: 5% methanol, 7% acetic acid

2.2.5 Mobility shift assays (gel retardation experiments)

Mobility shift assays were performed by a method based on Kulmburg *et al.,* (1992a).

2.2.5.1 Purifying DNA fragments: DNA fragments were purified from agarose gels or non-denaturing polyacrylamide gels before and/or after labelling.

Non-denaturing polyacrylamide gels (for 50mls): 7.5 ml 38% acrylamide/2% N,N'-methylene bisacrylamide, 5 ml 10 X TBE (Method I), 37.5 ml water, 50 µg ammonium persulfate and 20 µl TEMED.

2.2.5.2 Labelling reaction:

procedure (A): 50 ng DNA fragment resuspended in water, 15 µCi α-^{32}P-dATP, 2 µl 10 X L buffer, 1 µl each of 0.5 mM dGTP, 0.5 mM dTTP and 0.5 mM dCTP made up to a volume of 20 µl. 10 u of the Klenow fragment from *E. coli* DNA polymerase I was added and the solution incubated at room temperature for 30 min.

procedure (B): 20 µCi γ-^{32}P-dATP, 2 µl 10 X polynucleotide kinase buffer, 2 µl 10 mM spermidine, 10 u T4 polynucleotide kinase in a final volume of 20 µl was incubated at 37 °C for 1hr.

After labelling DNA fragments were purified using one of the following procedures:

procedure (A): 100 µl of 10 mM Tris.HCl pH 8.0, 0.1 mM EDTA (or water), was added. The labelled DNA fragment was purified by phenol/ $CHCl_3$ extraction and precipitated with 1/10th volume of 3 M NaAc and 2.5 X volume ethanol at

-20 °C for 15 min. DNA was pelleted by centrifugation at 12,000g for 30 min, rinsed in 70% ethanol and dried *in vacuo*. DNA was redissolved in 15 µl dH_2O. procedure (B): the reaction volume was increased to 30 µl with water and the mixture applied to a CL6B sepharose column.

2.2.5.3 Protein-DNA binding reaction: 1 µl (5 ng) of labelled fragment, 200 mM KCl, 2 µg poly{dI-dC} and 8 mM spermidine in a total of 10 µl was incubated at 25 °C with 10 µl of PBS, *E. coli* protein extract or nuclear protein extract for 15 min.

2.2.5.4 Electrophoresis Samples were then loaded onto a mobility shift gel and electrophoresed at 4 °C at 17 volts/cm in 0.25 X TBE (Method I), until the bromophenol blue tracker dye migrated 2/3 the way down the gel. Gels were wrapped in gladwrap, frozen (-80 °C) and exposed to film.

Mobility shift gels (for 40mls): 8 ml 50% glycerol, 8 ml 30% acrylamide/0.36% N,N'-methylene bisacrylamide, 1 ml 10 X TBE (Method I), 23 ml dH_2O, 33 µg ammonium persulfate, 20 µl TEMED.

2.2.6 DNase I sensitivity assays (footprinting)

DNase I sensitivity assays were based on the method of Kulmburg *et al.*, (1992a).

2.2.6.1 Labelling reaction: DNA fragments of 130-150 bp in length were labelled on one strand at one end only using procedure (A), section 2.2.5.2, except that 5 units of Sequenase^R enzyme was used instead of the Klenow fragment from *E. coli* DNA polymerase I and typically 2 X the amount of DNA fragment was used. The fragment was purified as described in 2.2.5.2 (procedure (A)).

2.2.6.2 Protein-DNA binding reactions were set up as described in 2.2.5.4 except that more labelled fragment was used, up to 10 ng. Duplicate samples were incubated as before then 2 µl of 50 mM $CaCl_2$, 100 mM $MgCl_2$ and 1 µl DNase I (in diluted in DNase I dilution buffer) was added for 1 min at 25 °C. The amount of DNase I was empirically

determined and two concentrations of 50 ng/µl and 100 ng/µl were used. Samples were loaded onto a mobility shift gel and electrophoresed as before (2.2.5.4).

2.2.6.3 Electrophoresis After electrophoresis and exposure to film, bands were excised and eluted with the crush and soak method of Sambrook *et al.*, (1989) with the addition of 10 mM 1,10-phenanthroline to the elution buffer or DNA was electrophoresed onto DEAE membranes and extraction was carried out overnight at 37 °C in 200-400 µl of 1 X TE, 1.5 M NaCl. DNA was recovered by phenol/ $CHCl_3$ extraction and ethanol precipitation. Duplicate DNA samples were combined, dissolved in 8 µl dH_2O and 12 µl of freshly prepared sequencing load buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol FF) and electrophoresed on a 10% sequencing gel (2,000 v until the bromophenol blue tracker dye migrated to the bottom of the gel) with 1 X TBE (Method II) as running buffer, fixed in 20% methanol/10% glacial acetic acid for 15 min, dried at 65 °C for 2 h and exposed to X-ray film.

10% Sequencing gel (for 40 ml): 7.5 ml 38% acrylamide/2% N,N'-methylene bisacrylamide, 21 g urea, 5 ml 10 X TBE (Method II), 40mg ammonium persulfate and 20 µl TEMED.

2.2.7 Maxam and Gilbert sequencing

The Maxam and Gilbert "G" reaction sequence was performed as described in Sambrook *et al.*, (1989). The Maxam and Gilbert "G+A" reaction sequence was performed as described by Muro *et al.*, (1993) using formamide that was not stored deionised.

2.2.8 Polymerase Chain Reaction (PCR) Methods

2.2.8.1 Conditions for making oligonucleotides double stranded Synthetic oligonucleotide 687 (2 µg, Table 2.3) was made double stranded using 2 µg of oligonucleotide 603 (Table 2.3), 2.5 mM $MgCl_2$, 0.125 mM dNTP's (N=AGTC), 15 µCi α-^{32}P-dATP, 3 u of Taq polymerase and buffer supplied by Bresatec. One drop of mineral oil (Sigma)

was added and the mixture incubated at 95 °C for 2 min followed by annealing/ extension at 60 °C for 1 min. The mixture was allowed to cool to room temperature slowly.

2.2.8.2 Amplification of oligonucleotides: 12 µl of resuspended DNA sample, 5 µCi α-^{32}P-dATP, 2.5 mM $MgCl_2$, 0.125 mM dNTPs (N=AGTC), 1 µg each of synthetic oligonucleotide 602 and 603, 1 X buffer supplied by Bresatec and 1 drop of mineral oil. The mixture was heated to 95 °C for 2 min and then 5 u of Taq polymerase was added. Conditions were 95 °C for 1 min, 60 °C for 1 min, for 30 cycles. After this the oligonucleotides were ethanol precipitated and resuspended in 12 µl of PBS.

2.2.8.3 Re-amplification of CASTing clones 1 µg of plasmid DNA sample, 5 µCi α-^{32}P-dATP, 2.5 mM $MgCl_2$, 0.125 mM dNTPs (N=AGTC), 0.125 µg each of synthetic oligonucleotide 602 and 603, 1 X buffer supplied by Bresatec and 1 drop of mineral oil. The mixture was heated to 95 °C for 2 min and then 5 u of Taq polymerase was added. Conditions were 95 °C for 30 sec, 45 °C for 30 sec, for 30 cycles. After this the oligonucleotides were isolated by electrophoresis on a non denaturing PAG (section 2.2.5.1).

2.2.9 Preparation of antibodies to glutathione-S-transferase(GST) fusion proteins

2.2.9.1 Antibodies raised against the GST-CREA (zinc finger region)

A New Zealand white rabbit was inoculated subcutaneously with a mixture of approximately 100-200 µg of column purified GST-CREA fusion protein with 44 µl of aluminium hydroxide as adjuvant (Imject^R alum; Pierce chemicals). The final concentration of aluminium hydroxide was 1.3 mg/100 µl. At 51 days the rabbit was inoculated with a further 600-1000 µg plus 264 µl adjuvant (final concentration of alum=1.3 mg / 100 µl). A test bleed was carried out at 71 days. 10 ml of blood was collected and a clot allowed to form. The serum was collected by centrifugation and the pellet discarded. The presence of anti-GST-CREA fusion protein antibodies was determined with Ochterlony plates (Sambrook *et al.*, 1989).

A final inoculation of 300-600 µg of fusion protein with 132 µl of Imject^R alum was performed at 89 days. At 103 days approximately 40 ml of blood was collected, allowed to clot and the serum collected by centrifugation (12,000g for 10 min). The serum was stored in 1 ml aliquots at -80 °C.

2.2.9.2 Antibodies raised against the GST-CREA(RGR) (carboxy terminal region)

Antibodies to the GST-CREA(RGR) fusion protein were raised in the same way as those of the zinc finger fusion protein except that the GST-CREA(RGR) protein was used and two rabbits were inoculated.

2.2.10 Western analysis

2.2.10.1 Western transfer conditions

The SDS polyacrylamide gel was rinsed in transfer solution (50 mM Tris, 40 mM glycine, 20-25% methanol) for one to three minutes and then assembled on an electroblotting apparatus with Immobilon PVDF (Millipore) as the recipient membrane. For transfers submerged in transfer solution using a Trans-Blot™ apparatus (Bio-Rad) a current density of 50 mA was applied for 2-4 h which was increased to 200 mA for at least 16 h. Semi-dry transfers were carried out using transfer solution and a Trans-Blot™SD apparatus (Bio-Rad) according to the manufacturers instructions.

2.2.10.2 Immunodetection

Membranes were rehydrated if necessary in 100% methanol prior to blocking by gentle agitation in 1 ml of blocking buffer (1XPBS, 1% skim milk powder, 0.1% Tween 20) per cm^2 of membrane for 0.5 h. To 5-7 ml of blocking buffer was added the antibody solution and gentle agitation was continued for 1-2 h. The membrane was washed 3 times in 1XPBS for a total of 20 min and then agitated for 1 h with 5-7 ml of blocking buffer containing 2 µl horse radish peroxidase (Promega). The membrane was washed 3 times for 20 min in 1XPBS and then drained well. 10 ml of visualisation solution (20% methanol, 5 mg 4-chloronapthol, 10 mM Tris.HCl pH 8.0, 150 mM NaCl, 0.1% H_2O_2) was added and the colorimetric reaction was stopped after 0.5 h by washing in 1XPBS.

2.2.11 Separation and Purification of antibodies

2.2.11.1 Using glutathione sepharose and the zinc finger antibodies

Separation and purification of antibodies was based on the protocol given in Harlow and Lane (1988). 1 ml of glutathione sepharose 4B beads (Pharmacia) was placed inside a 5 ml syringe with Whatman glass microfibre filter as support. All solutions were filtered through a 0.2 µm filter before loading onto the column. The column was washed with 10 ml of PBS and equilibrated with 10 ml of PBS+ 1% triton X-100.

0.5 ml of *E. coli* total protein extract derived from cells harbouring the pGEX-2T plasmid was diluted in 10 ml of PBS and then applied to the column five times and the final filtrate discarded. The column was rinsed with 10 ml of PBS. 1 ml of whole rabbit serum (diluted in 10 ml PBS) was applied to the column five times. The filtrate (A) was kept. Another column was prepared in the same way except that the protein extract attached to the column was derived from cells harbouring the pGEX-*creA* construct which produces the GST-CREA fusion protein. The above filtrate (A) was passed through this second column five times.

The anti-CREA antibodies attached by acid sensitive interactions were eluted with 10 ml 100 mM glycine and the filtrate collected in 1 ml of 1 M Tris.HCl pH 8.0. The column was washed with 10 mM Tris.HCl until the pH rose to 8.8 (approximately 20 ml). The anti CREA antibodies attached by base sensitive interactions were then eluted with 10 ml of 100 mM triethylamine, prepared fresh and the filtrate collected in 1 ml of 1 M Tris.HCl pH 8.0. The antibody fractions were kept separately and were dialyzed against PBS, precipitated with an equal volume of saturated ammonium sulfate and then re-dialyzed. Antibodies were resuspended in 300 µl of PBS.

2.2.11.2 Using glutathione sepharose and the RGR1 antibodies

5 ml of glutathione sepharose 4B beads (Pharmacia) was placed inside a 5 ml syringe with Whatman glass microfibre filter as support. All solutions were filtered through a 0.2 µm filter

before loading onto the column. The column was washed with 20 ml of PBS and equilibrated with 5 ml of PBS+ 1% triton X-100.

0.5 ml of *E. coli* total protein extract derived from cells harbouring the pGEX-2T plasmid was applied to the column and the filtrate discarded. The column was rinsed with 30 ml of PBS. 0.5 ml of rabbit serum was applied to the column and the column was washed with 8 ml of PBS. The eluates were combined and 1% skim milk powder, 0.85 ml of 10XPBS was added before this antibody solution was used in immunodetections (section 2.2.10.2).

2.2.11.3 Using protein A-sepharose columns (both sets of antibodies)

A 4 ml protein A-sepharose column was made, to which 5 ml of serum with 0.5 ml 1 M Tris.HCl pH 8.0 was applied. The column was washed twice with 40 ml of Tris.HCl pH 8.0 and then the antibodies were removed from the column with 16 ml 100 mM glycine pH 3.0. The 32 fractions of 0.5 ml were collected in 50 µl 1 M Tris.HCl pH 8.0 and concentrated using centricon 10's (Amicon).

2.2.12 Isolation of *A. nidulans* protein extracts

2.2.12.1 Nuclear extracts

Nuclear extracts were prepared based on the procedure given in Richardson *et al.*, (1992). Approximately 10^9 conidia were inoculated into 800 ml *Aspergillus* nitrogen free medium and grown at 37 °C for 16 h. Mycelium was harvested, washed with cold distilled H_2O, and blotted dry. Mycelium was then immediately ground under liquid nitrogen to a fine powder and added to 15 ml of buffer A (1 M sorbitol, 7% ficoll, 20% glycerol, 5 mM MgAc, 5 mM EGTA, 3 mM $CaCl_2$, 50 mM Tris.HCl (pH 7.5), 5 mM DTT, 18 mM PMSF). After 10 min on ice with occasional stirring. 24 ml of buffer B (10% glycerol, 25 mM Tris.HCl (pH 7.5), 5 mM MgAc, 5 mM EGTA,1 mM DTT, 2.5 mM PMSF) was added slowly and the mixture was layered over 2.5 ml of 2 X buffers A and B in a ratio of 2.5:4. Mycelial debris was pelleted by spinning at 1,500g for 7 min in a swing out rotor. The supernatant was then layered over 10 ml of 2 X

buffer C (25 mM Tris.HCl (pH 7.5), 10% glycerol, 1 M sucrose, 5 mM MgAc, 3 mM DTT) and spun at 8,000g for 25 min. The nuclear pellet was suspended in 1 ml of nuclear extraction buffer (15 mM HEPES-KOH (pH 7.5), 0.1 mM EGTA, 5 mM $MgCl_2$, 5% glycerol, 0.4 M NaCl, 1 mM DTT, 1 mM PMSF, 12.5 µg of pepstatin A per ml) and rocked gently on ice for 30 min. The mixture was spun at 86,000g at 2 °C for 30 min to pellet chromosomal DNA. The supernatant was dialyzed against 400 ml of binding buffer (25 mM HEPES-KOH (pH 7.6), 40 mM KCl, 1 mM EDTA, 10% glycerol, 1 mM DTT, 2 mM PMSF) at 4°C overnight. The protein extract was concentrated by centrifugation in a Centricon 10 tube for 2 to 4 h. The soluble protein concentration in the final extract was measured by the method of Bradford (1976), using commercially available reagents (Bio-Rad laboratories).

2.2.12.2 Total protein extracts

Approximately 10^7 conidia were inoculated into 50 ml *Aspergillus* nitrogen free medium or Carbon free medium with 1% arabinose as carbon source and grown at 37 °C for 16 h. Mycelia were harvested, washed with cold distilled H_2O, blotted dry and then immediately ground with sand to a fine paste. Samples were centrifuged at 12,000g for 5 mins and the supernatant frozen at -80°C in small aliquots.

2.2.13 Isolation of nucleic acids from *A. nidulans*

DNA was isolated by the method of Hynes *et al.*, (1983), from freeze-dried *A. nidulans* mycelia harvested from cultures grown overnight in glucose medium.

For the isolation of *A. nidulans* RNA, cultures were grown overnight (37 °C) in glucose media, harvested, washed with a large volume of prewarmed carbon free liquid medium, and transferred to fresh prewarmed glucose medium for three to six hours (37 °C). These cultures were then harvested, washed with a large volume of sterile distilled water and freeze-dried overnight. Total RNA was then extracted by the method of Reinert *et al.*, (1981).

2.2.14 *in vivo* Labelling of *A. nidulans* total proteins with $^{35}SO_4^{2-}$

Total proteins were radio-labelled by the method of Lockington (1984). To 50 ml of low sulfate minimal medium was added 10 mM ammonium tartrate plus any other supplements required by the strain to grow. 0.5 μCi of $^{35}SO_4^{2-}$ was added and the media inoculated with 1.6 X 10^8 conidia. Cultures were left to grow at 37 °C overnight and were then harvested by filtration through Whatman filter paper (No. 4) and pressed dry between paper towel. Labelled mycelia were stored at -80 °C or used immediately.

Proteins were extracted by grinding with glass, or acid purified sand, in ice cold immunoprecipitation buffer. Mycelial debris was pelleted by centrifugation for 5 min at 12,000g. The supernatant was removed and diluted in 10 mM Tris.HCl, pH 8.0 for control tracks using SDS PAGE or immunoprecipitated as below.

2.2.15 Immunoprecipitation of proteins

Proteins were immunoprecipitated using a method based on the procedure by Thorley-Lawson (1979). 150 μl of protein A-sepharose mixture (equivalent to 150 μg protein A) was washed in immunoprecipitation buffer and resuspended in a final 130 μl of immunoprecipitation buffer. 500 μl of protein extract was incubated on ice for 30 min with 3 μl of antibody containing serum (from the final bleed), or 20 μl of acid/base purified CREA antibodies. The 130 μl of beads were then added and this mixture gently agitated on ice for a further hour. The beads were spun briefly and rinsed twice with 1 ml of ice cold 1% triton X-100, 0.5 M LiCl, 10 mM Tris.HCl, pH 8.0. Twice with 1 ml of ice cold 1% triton X-100, 10 mM Tris.HCl, pH 8.0 and three times with 1 ml of ice cold 10 mM Tris.HCl, pH 8.0. The beads were resuspended in 40 μl of gel loading buffer (10% glycerol, 3% SDS, 60 mM Tris.HCl (pH 6.8), 1 M β-mercaptoethanol, 0.1% bromophenol blue) and boiled for 2 min.

CHAPTER 3

Chapter 3

In Vitro DNA Binding Studies Using a CREA Fusion Protein

A number of techniques have been used to detect binding by proteins to DNA. The gel mobility shift assay (discussed in detail by Lane *et al.*, 1992) was first developed by Garner and Revzin (1981) and is a simple method using electrophoresis to separate DNA complexed with proteins from unbound DNA. The different mobilities of DNA and DNA-protein complexes can be visualised either by ethidium bromide staining, radiolabelling of the DNA or by using labelled protein extracts. In most cases DNA bound to proteins has a reduced mobility during electrophoresis, thus this procedure is often called a gel retardation analysis.

A more detailed indication of the base pairs protected by a DNA binding protein can be obtained by DNase I sensitivity assays (Galas and Schmitz, 1978). In this technique a labelled DNA fragment is incubated with the protein of interest, partially digested with DNase I and the products resolved on a sequencing gel. A region, or window, which is free from bands corresponds to the bases in the DNA fragment that were protected by bound protein. Chemical agents which can be used in the process to cleave DNA, for example dimethyl sulfate (DMS), can also be used in a similar manner to identify specific bases that are important in the interactions with the protein (reviewed by Saluz and Jost, 1993).

In this chapter experiments are described in which an *E. coli* expressed CREA fusion protein was used to investigate the binding of CREA to the 5' regions of two genes which are subject to carbon catabolite repression: *amdS*, the structural gene for acetamidase, and *facB,* a regulatory gene involved in acetate utilisation (Hynes and Kelly, 1977; Kelly and Hynes, 1977; Hynes *et al.*, 1983; Katz and Hynes, 1989b). To investigate the possibility of an autoregulatory role for the *creA* gene product (Dowzer and Kelly, 1991), binding to the 5' region of *creA* was also undertaken. The aim of these experiments was to determine whether CREA bound to DNA *in vitro*, in a sequence specific manner, using the 5' regions of genes where evidence existed for *creA* control.

3.1 Isolation of an *E. coli* expressed fusion protein between GST (glutathione-S-transferase) and CREA

To test the hypothesis that CREA was a DNA binding protein it was first necessary to obtain sufficient quantities of CREA protein. One approach was to express CREA as a fusion protein in *E. coli*. Many researchers have demonstrated DNA binding of cloned gene products by expressing them as fusion proteins (Alex *et al.*, 1992; Suarez *et al.*, 1995) and other researchers have used small peptides expressed in *E. coli* for X-ray crystallographic analysis of DNA binding (Kissinger *et al.*, 1990; Pavletich and Pabo, 1991).

The vector pGEX-2T, which encodes the glutathione-S-transferase gene (GST, 26 kDa) from the parasitic helminth *Schistosoma japonicum* (Smith and Johnson, 1988), was chosen since GST fusion proteins have been shown to be stable and often soluble (Smith and Johnson, 1988) and are able to be purified using glutathione sepharose columns (Pharmacia). The construct used to express CREA as a fusion protein (Fig. 3.1a) contains the two putative zinc fingers and the alanine rich region of CREA (Kulmburg *et al.*, 1993).

E. coli total protein extracts were isolated as described in section 2.2.2. Purification by affinity column chromatography using a glutathione-sepharose column (section 2.2.3) and subsequent analysis by SDS PAGE (section 2.2.4) was carried out to assess the amount and size of the GST-containing proteins. From pGEX-2T transformed cells a protein of 26 kDa, corresponding to the region of GST coded by the pGEX-2T plasmid, was produced (Fig. 3.1b). Using the method of Bradford (1976), the yield was determined to be 30 mg of total soluble protein per 100 ml of bacterial culture. It was estimated that 30% of the total protein extract consisted of the 26 kDa GST protein. Therefore the concentration of the GST protein was estimated to be 10 mg/ ml. After column purification very few other proteins were visible.

A full-length GST-CREA fusion protein of 52 kDa was produced from cells transformed with a recombinant pGEX-*creA* plasmid (Fig. 3.1b). As well as this full length protein, smaller

column purified protein products were present. Since the B strain BL21 cells used to produce these foreign proteins should lack the La protease (Studier *et al.*, 1990), these were probably the result of premature termination or degradation by other proteases. The yield was determined to be 16 mg of total soluble protein per 100 ml of bacterial culture. The percentage of full length fusion protein in the total *E. coli* extracts was estimated to be approximately 5-10% giving a concentration for the fusion protein of 0.8-1.6 mg/ ml. After column purification a 70 kD protein was seen. It has been suggested[1] that this protein is likely to be the chaperonin dnaK, which often associates with foreign proteins in *E. coli* to bring about proteolytic degradation of foreign proteins (Sherman and Goldberg, 1992). The 70 kDa dnaK contaminant was present in the column purified GST Track (Track 2, Fig. 3.1b) but at a much reduced level .

3.2 Mobility shift analysis of the 5' *amdS* region

The GST-CREA fusion protein was used in gel mobility shift assays to determine whether the fusion protein produced by *E. coli* was capable of sequence specific DNA-binding. For this purpose, the 5' region of *amdS* was chosen since molecular genetic studies have shown that *amdS* is under *creA* control (discussed in section 1.3.1.5). A map of the 5' region of *amdS* showing the restriction endonuclease sites used in this study is shown in Figure 3.2. Several fragments from this region were sub-cloned to allow sufficient material for further study. A diagrammatic representation of these clones is also shown in Figure 3.2.

3.2.1 Mobility shift analysis of the BP343 region

The *Pst* I-*Bam* HI region from -215 to +128, relative to the start point of translation, was analysed for binding to CREA. The insert from pBP343 was excised by digestion with the restriction endonucleases *Bam* HI and *Hind* III and purified by electrophoresis on an agarose gel for use in a gel mobility shift assay (section 2.2.5). Unless otherwise stated 5 pmol of DNA fragment was used in gel mobility shift assays and experiments were always repeated at least

[1]Trouble shooting guide from the "GST Gene Fusion System" mannual, copyright © (1993), Pharmacia P-L Biochemicals Inc,

Figure 3.1a The construct between the vector pGEX-2T and the cloned *creA* gene

The *Nco* I-*Nco* I fragment from pCD5 (Table 2.2) is present in the *Sma* I site of the expression vector pGEX-2T (Kulmburg *et al.*, 1993). The vector pGEX-2T encodes the glutathione-S-transferase (GST) gene under the control of the *tac* promoter and contains stop codons in all three frames (Smith and Johnson, 1988). The GST-CREA fusion protein contains 206 amino acids from CREA (out of 416 amino acids) including the zinc-finger region and the alanine rich region.

Figure 3.1b Electrophoretic analysis of the fusion protein

A 10 % SDS-polyacrylamide gel (section 2.2.4) stained with Coomassie Brilliant Blue^{R}.

Track 1 80 µg of total protein extract from *E.coli* (section 2.2.2) transformed with the non-recombinant plasmid pGEX-2T.

Track 2 10 µg of the glutathione-sepharose column purified (section 2.2.3) protein extract from *E.coli* harbouring the non-recombinant plasmid pGEX-2T. The arrow shows the position of the purified GST protein.

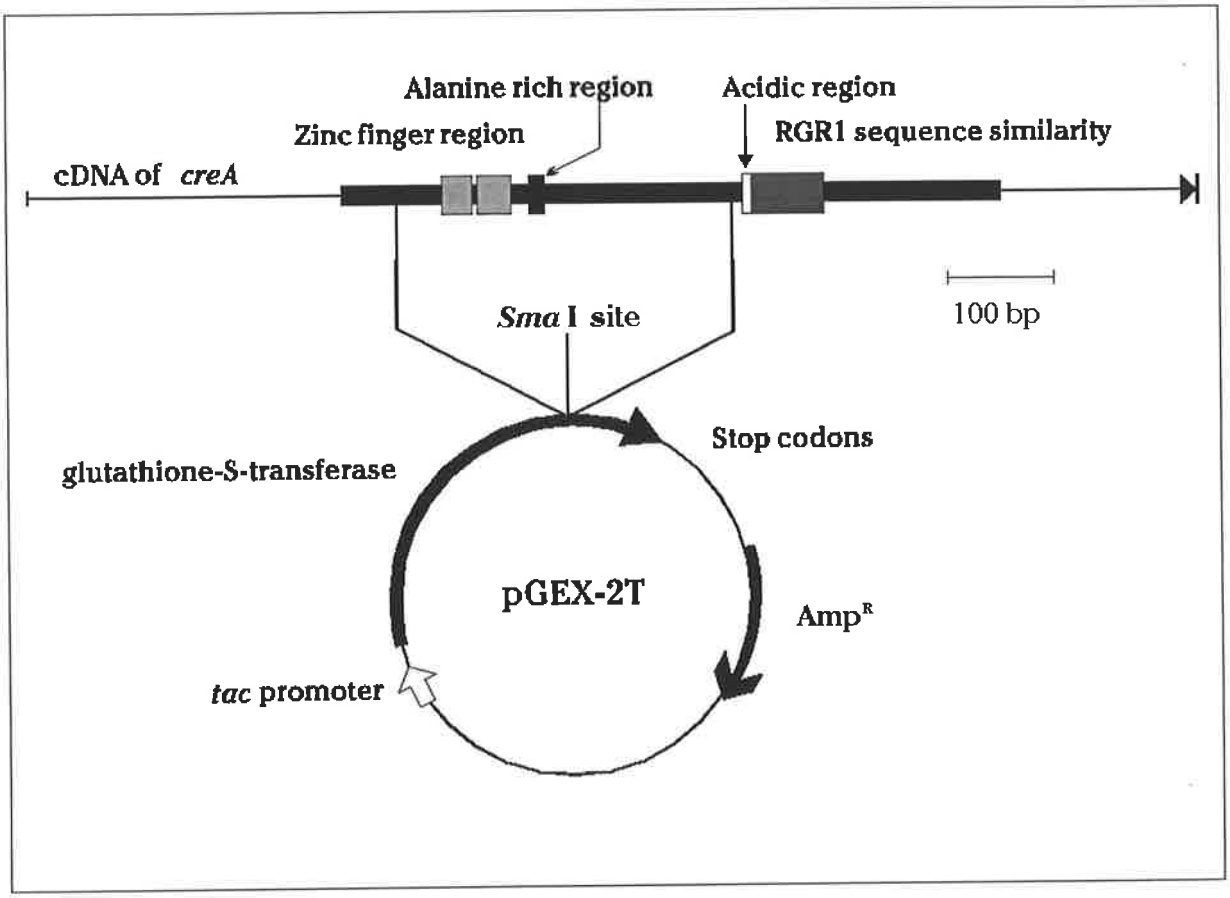
Track 3 1µg prestained molecular weight markers with the following (apparent) molecular weights:

Protein	Size (kDa)
α2-macroglobulin	190,000
β-galactosidase	125,000
fructose-6-phosphate kinase	88,000
pyruvate kinase	65,000
fumarase	56,000
lactate dehydrogenase	38,000
triosephosphate isomerase	33,500

Track 4 150 µg total *E.coli* protein extract from cells transformed with the recombinant pGEX-*creA* construct.

Track 5 13 µg of the glutathione-sepharose column purified protein extract from *E.coli* harbouring the recombinant pGEX-*creA* construct. The lower arrow shows the position of the full length GST-CREA protein and the upper arrow the 70 kDa protein which co-column purifies.

a)


Alanine rich region
Acidic region
Zinc finger region
RGR1 sequence similarity
cDNA of creA
100 bp
Sma I site
Stop codons
glutathione-S-transferase
pGEX-2T
Amp^R
tac promoter

b)

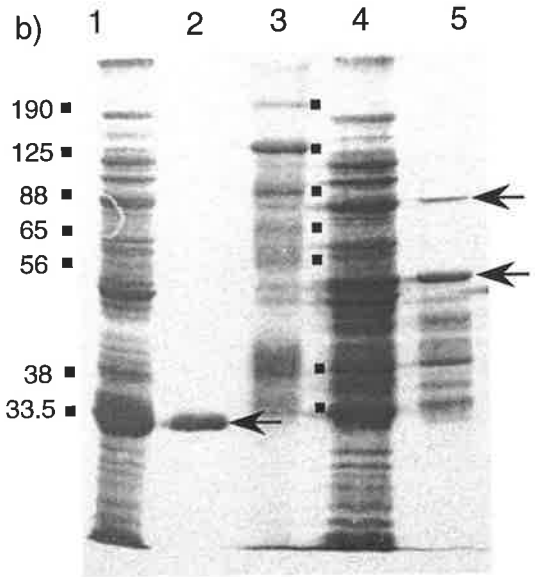

1
2
3
4
5
190
125
88
65
56
38
33.5

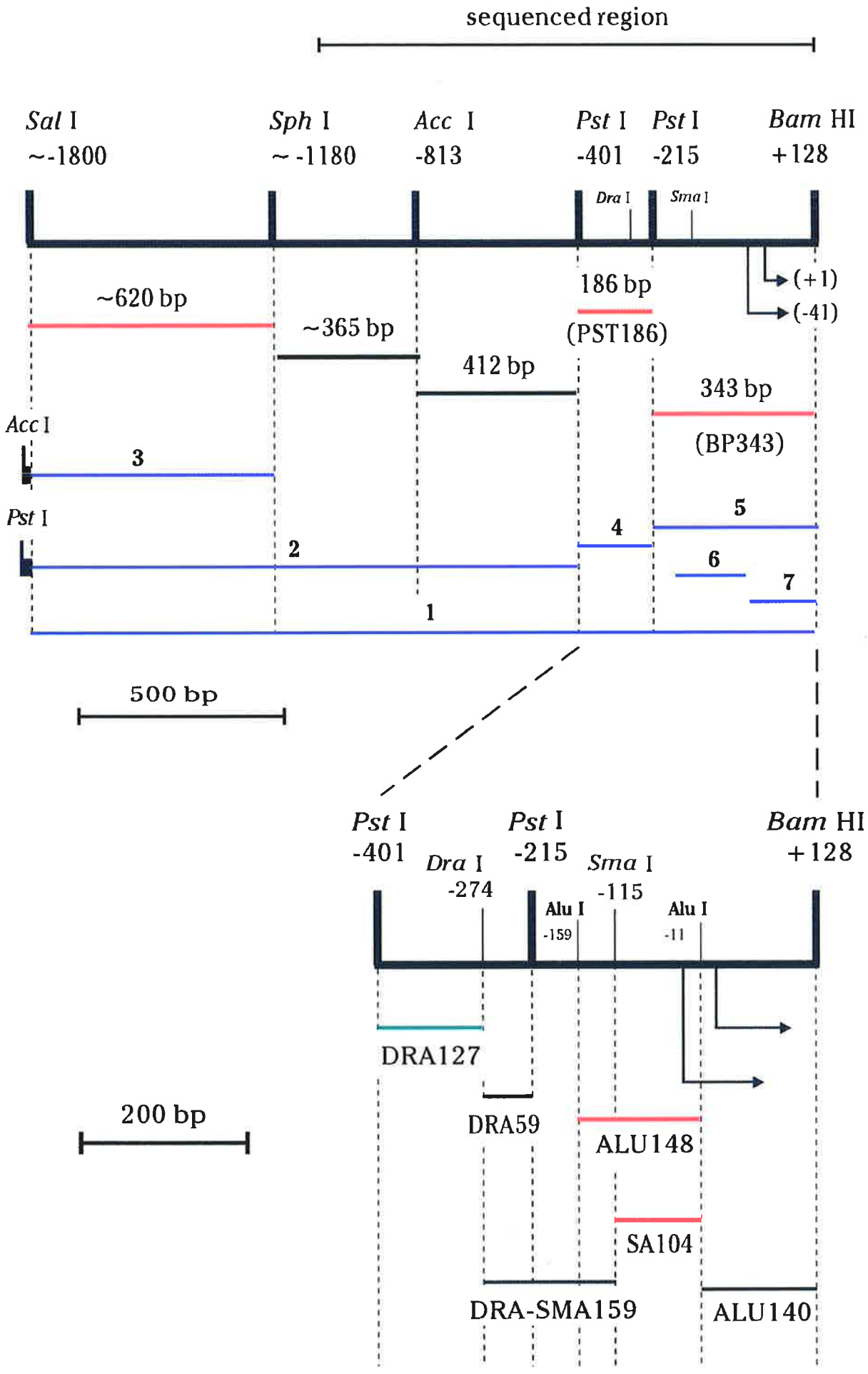
Figure 3.2 Partial restriction endonuclease map of the 5' region of *amdS*

Scale map of the 5' region of the *amdS* gene (Corrick *et al.*, 1987). Fragments isolated for mobility shift assays are shown as horizontal bars with sizes (excluding polylinker regions from sub-clones) marked above the bars. Fragments are referred to in the text by size and by the names of the restriction endonucleases used to isolate them. Fragments which were found to bind GST-CREA strongly are shown in red and one with medium affinity is shown in green. As indicated the start point of translation is +1.

Blue horizontal lines indicate the clones constructed from the 5' region of the *amdS* gene. Numbers above these lines refer to the Table below. The black segments represent regions of polylinker with restriction endonuclease sites used in the cloning procedure and/or referred to in the text.

Plasmid clones constructed from the 5' region of the *amdS* gene

Number	Name	Cloning Description
1	pBAMSAL1.8	1.8 kb *Sal* I-*Bam* HI fragment from p3SR2 (Table 2.2) Cloned into *Bam* HI and *Sal* I digested pUC19
2	pPSP1.2	1.2 kb *Pst* I fragment from pBAMSAL1.8 cloned into *Pst* I digested pUC19
3	pAMDS660	*Acc* I-*Sph* I fragment of approx. 660 bp from pPSP1.2 cloned into *Acc* I and *Sph* I digested pUC19
4	pPST186	186 bp *Pst* I fragment cloned into *Pst* I digested pUC19
5	pBP343	*Pst* I-*Bam* HI fragment of 343 bp cloned into *Pst* I and *Bam* HI digested pUC19
6	pALU148	*Alu* I fragment of 148 bp cloned into the *Sma* I site of pUC19
7	pALU140	140 bp *Alu* I-*Bam* HI fragment cloned into pUC19 digested with *Sma* I and *Bam* HI


sequenced region
Sal I
~-1800
Sph I
~-1180
Acc I
-813
Pst I
-401
Pst I
-215
Bam HI
+128
Dra I
Sma I
(+1)
(-41)
~620 bp
~365 bp
412 bp
186 bp
(PST186)
343 bp
(BP343)
Acc I
Pst I
1
2
3
4
5
6
7
500 bp
Pst I
-401
Pst I
-215
Bam HI
+128
Dra I
-274
Sma I
-115
Alu I
-159
Alu I
-11
DRA127
DRA59
ALU148
SA104
DRA-SMA159
ALU140
200 bp

once. The BP343 fragment was radiolabelled using the Klenow fragment of DNA polymerase I as described in section 2.2.5.2, and incubated in binding buffer, with either PBS, total *E. coli* protein extract producing GST or total *E. coli* protein extract producing the GST-CREA fusion protein (section 2.2.5.3). The mixture was immediately electrophoresed on a 6% polyacrylamide gel (section 2.2.5.4). As shown in Figure 3.3a, the mobility of the BP343 fragment was greatly reduced by the presence of the GST-CREA fusion protein but not by the presence of the GST protein extract. These results show that the CREA fusion protein was able to bind to DNA, and this binding required no other *A. nidulans* proteins. The GST-CREA protein extract contained GST-CREA proteins of varying size (Fig. 3.1b) and some of these may be able to bind DNA. This may explain why the retarded complex was not a discrete band.

To ensure that the GST-CREA fusion protein had bound to DNA in a sequence specific manner, a control fragment of 480 bp isolated from bacteriophage *SPP*1 digested with *Eco* RI was labelled and incubated as previously described for BP343. As shown (Fig. 3.3a) there was no retarded band although some generalised smearing back of the SPP1-480 fragment in the presence of the GST-CREA fusion protein was seen.

The BP343 fragment is quite large so in order to narrow down the binding region BP343 was digested with *Alu* I restriction endonuclease and the two fragments ALU148 and ALU140 (Fig. 3.2) were sub-cloned. These were tested for the ability to bind CREA and the results are shown in Figure 3.3b. The ALU148 fragment showed very clear retardation of mobility in the presence of the fusion protein as well as some retardation in the presence of the GST containing *E. coli* extracts. However, the complex formed with the GST containing *E. coli* extract migrated faster than that formed with the CREA fusion protein. Binding by the GST containing extract was not always detected, and when present the complexes always migrated faster. The CREA fusion protein was able to retain DNA binding ability after freezing and thawing, whereas the pGEX-2T extract lost DNA binding activity after as few as 2 freeze thaw cycles. Thus this may explain why binding by the pGEX-2T extract was not always seen. The GST-CREA fusion protein was

Figure 3.3a Mobility shift analysis of the BP343 and PST186 fragments

A mobility shift assay (section 2.2.5) of BP343 (**Tracks 1-3**) and PST186 (**Tracks 4-6**) from the 5' region of *amdS* and the 480bp *Eco* RI fragment (SPP1-480) from the bacteriophage SPP-1 (**Tracks 7-9**). Labelled DNA fragments were incubated with 10 µl of PBS (**Tracks 1, 4 & 7**) or 300 µg total protein extract from pGEX-2T transformed *E.coli* producing GST (**Tracks 2, 5 & 8**) or 160 µg total protein extract from pGEX-*creA* transformed *E.coli* producing GST-CREA (**Tracks 3, 6 & 9**).

w= wells

Figure 3.3b Further mobility shift analysis of the BP343 fragment

A mobility shift assay of ALU148 (**Tracks 1-3**) and ALU140 (**Tracks 4-6**) from *amdS*. Labelled DNA fragments were incubated with PBS (**Tracks 1 & 4**) or 150 µg total protein extract containing GST (**Tracks 2 & 5**) or 80 µg total protein extract containing GST-CREA (**Tracks 3 & 6**).

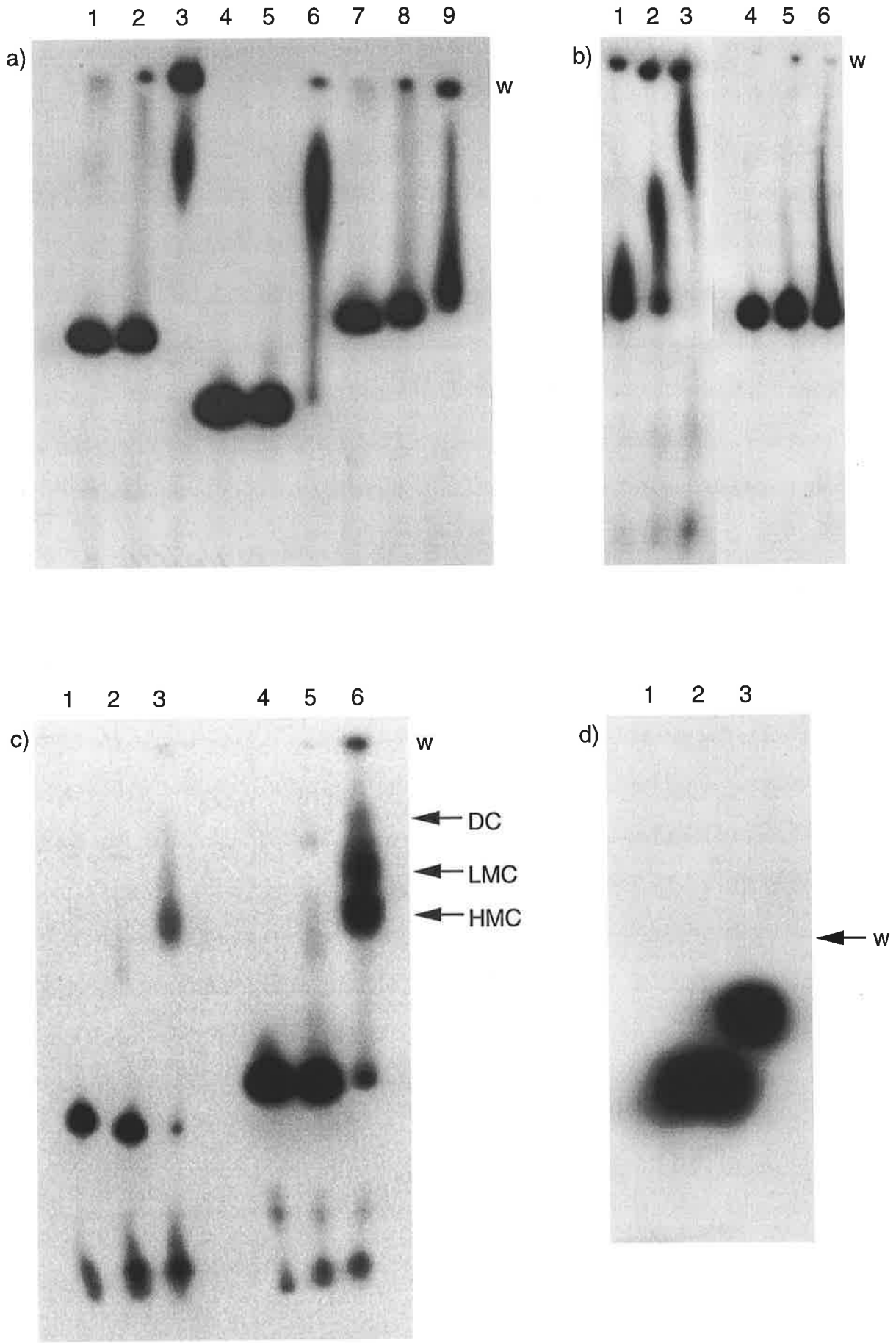
Figure 3.3c Mobility shift analysis of the SA104 sub-fragment of ALU148

Comparison of mobility shift assays using radiolabelled SA104 (**Tracks 1-3**) and ALU148 (**Tracks 4-6**). Labelled DNA fragments were incubated with PBS (**Tracks 1 & 4**) or 300 µg total protein extract containing GST (**Tracks 2 & 5**) or 160 µg total protein extract containing GST-CREA (**Tracks 3 & 6**).

HMC= higher mobility complex
LMC= lower mobility complex
DC= dimerisation complex (discussed in chapter 4)

Figure 3.3d Mobility shift analysis using agarose gel electrophoresis

A mobility shift assay using radiolabelled ALU148 fragment was carried out as described in section 2.2.5 except that an agarose gel was loaded instead of a polyacrylamide gel. Labelled ALU148 fragment was incubated with PBS (**Track 1**), 300 µg total protein extract containing GST (**Track 2**) or 160 µg total protein extract containing GST-CREA (**Track 3**). The position of the wells is indicated by the arrow. Shown is a long exposure.


a)
1 2 3 4 5 6 7 8 9
w
b)
1 2 3 4 5 6
w
c)
1 2 3 4 5 6
w
DC
LMC
HMC
d)
1 2 3
w

able to bind the ALU148 fragment even in the presence of these other *E. coli* protein(s). It may be that CREA has a higher affinity for this DNA fragment or that both proteins are able to bind simultaneously. The ALU140 fragment showed no altered mobility in the presence of the CREA fusion protein (Fig. 3.3b).

A sub-fragment *Sma* I-*Alu* I (SA104) of ALU148 was tested to see if it was able to bind GST-CREA as effectively as ALU148. For this, pALU148 was digested with *Sma* I and *Bam* HI and the labelled fragment was used in a gel retardation experiment (Fig. 3.3c). SA104 was able to bind CREA. On higher resolution gels, the retardation of ALU148 resolves into a higher (HMC) and lower (LMC) mobility complexes but for SA104 there was very little evidence of formation of a lower mobility complex. As similar amounts of DNA and the same protein extract was used for both fragments, this difference must be caused by differences in their sequences.

3.2.2 Analysis of the interaction between the CREA fusion protein and the ALU148 fragment from *amdS*

3.2.2.1 Are any fusion protein-DNA complexes migrating toward the cathode?

It was possible that some complexes were migrating toward the cathode and therefore would not be detected using vertical polyacrylamide gels. To determine whether there was any migration toward the cathode a mobility shift assay was prepared in the usual way (section 2.2.5) except that a 0.7% horizontal agarose gel, with the wells cast in the centre, was loaded instead of a polyacrylamide gel. Any migration of the complexes to either the anode or the cathode could be detected. As shown in Figure 3.3d, no labelled DNA fragments migrated toward the cathode. The reduction in radioactivity observed in Tracks 1 and 2 compared to Track 3 is caused by the diffusion of smaller uncomplexed DNA fragments out of the agarose gel.

3.2.2.2 Is the fusion protein in excess?

To determine whether the fusion protein was in excess compared to the DNA fragment, a series of protein dilutions were used in gel mobility shift assays (Fig. 3.4a). From these experiments

it was clear that two electrophoretically distinct complexes (HMC and LMC) were formed by the fusion protein extract with this DNA fragment. This may be due to the presence of two sites, where one site, being of higher affinity is occupied first, and the other of lower affinity requires a higher concentration of fusion protein extract for occupancy. On the other hand the lower mobility complex may be due to the dimerisation of the fusion protein through the GST moieties at higher concentrations. Dimerisation of native GST is known to occur *in vivo* (Mannervik, 1985). The fusion protein needed to be in excess in order to force the majority of the labelled DNA fragment into the lower mobility complex. The addition of 16 µg of GST-CREA containing *E. coli* extract (1/10th the amount used in previous experiments) still produced almost total retardation of the ALU148 fragment.

3.2.2.3 Is binding to the labelled fragment able to be inhibited by the addition of cold DNA fragments?

Using cold ALU148 fragment as a competitor only very slight competition was visible for 80-160 µg amounts of protein. Using a smaller amount of protein (0.16 µg), a 100 fold excess of cold ALU148 fragment completely inhibited binding to the labelled DNA (Fig. 3.4b). Since a non-competing fragment was not used,there is no evidence that this competition was specific.

The cold SA104 fragment was used in competition experiments (Fig. 3.4c) and was found to compete with the ALU148 fragment. Furthermore it competitively inhibited the formation of both electrophoretically distinct complexes. Whether it was able to do this by lowering the effective unbound concentration of the fusion protein or by directly competing with the ALU148 fragment for the binding of the low mobility complex, is not known.

3.2.2.4 Mobility shift assays using column purified protein extracts

Since the fusion protein could be easily purified from *E. coli* proteins using affinity chromatography, glutathione-sepharose column purified CREA fusion protein was used in gel mobility shift assays. The ALU140 fragment which was shown not to bind the total *E. coli* protein extract containing GST-CREA and the ALU148 fragment were tested. The complex

Figure 3.4a Mobility shift analysis with decreasing amounts of fusion protein

A mobility shift assay using radiolabelled ALU148 fragment was performed with decreasing amounts of GST-CREA fusion protein. Tracks contain labelled ALU148 fragment incubated with:

Track 1 PBS
Track 2 300 μg total protein extract containing GST
Track 3 160 μg total protein extract containing GST-CREA
Track 4 80 μg total protein extract containing GST-CREA
Track 5 16 μg total protein extract containing GST-CREA
Track 6 1.6 μg total protein extract containing GST-CREA
Track 7 0.16 μg total protein extract containing GST-CREA

Figure 3.4b Competition with cold ALU148 fragment

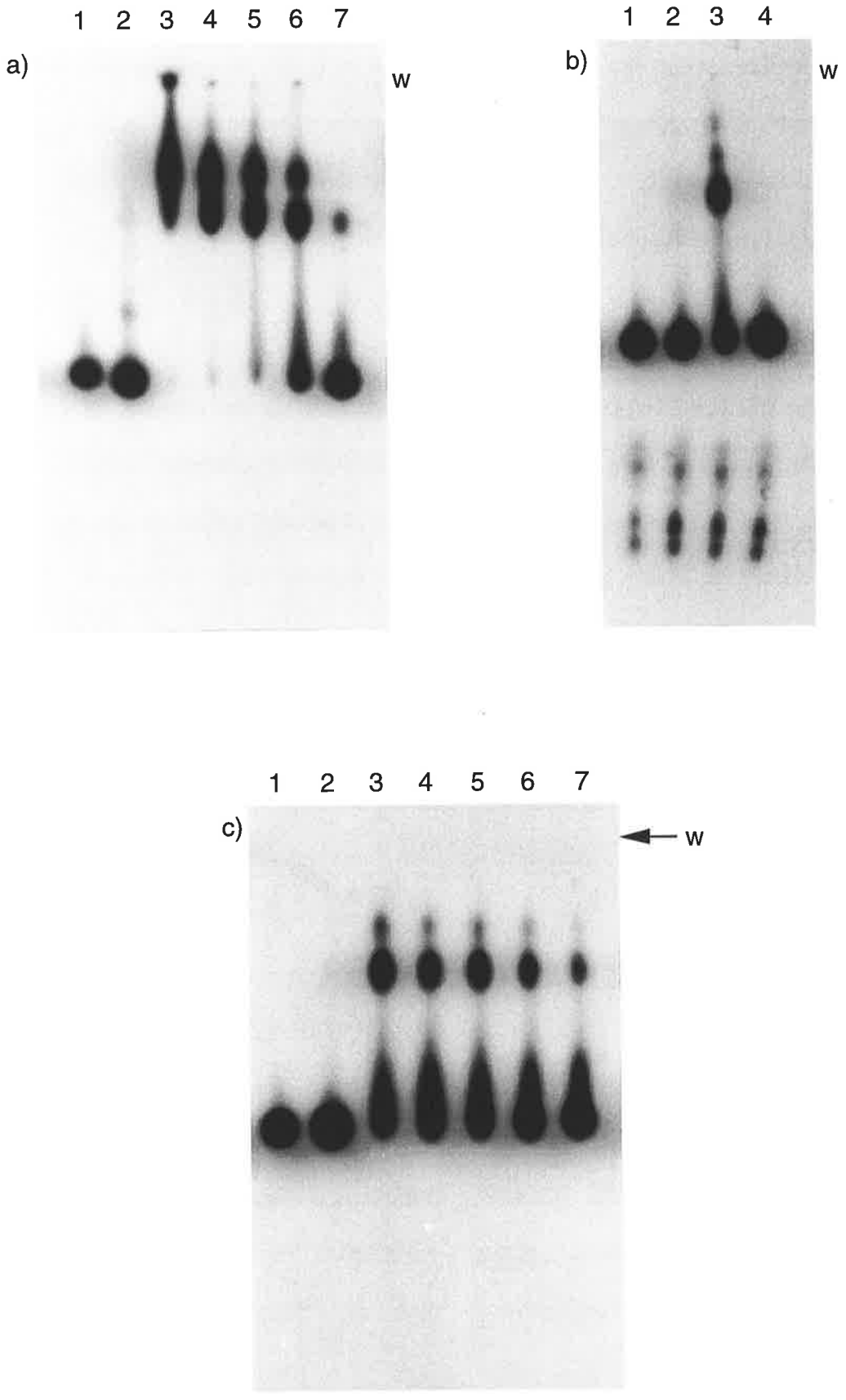
A mobility shift assay. Tracks contain 5 ng labelled ALU148 fragment incubated with:

Track 1 PBS
Track 2 3 μg total protein extract containing GST
Track 3 0.16 μg total protein extract containing GST-CREA
Track 4 0.16 μg total protein extract containing GST-CREA with 500 ng cold ALU148

Figure 3.4c A competition experiment with the SA104 sub-fragment of ALU148

5 ng of radiolabelled ALU148 fragment per Track was incubated with increasing amounts of cold SA104 DNA. Tracks contain:

Track 1 PBS
Track 2 3 μg total protein extract containing GST
Track 3 0.16 μg total protein extract containing GST-CREA
Track 4 0.16 μg total protein extract containing GST-CREA with 5 ng cold SA104
Track 5 0.16 μg total protein extract containing GST-CREA with 10 ng cold SA104
Track 6 0.16 μg total protein extract containing GST-CREA with 50 ng cold SA104
Track 7 0.16 μg total protein extract containing GST-CREA with 200 ng cold SA104


1 2 3 4 5 6 7
a)
w
1 2 3 4
b)
w
1 2 3 4 5 6 7
c)
w

produced with both fragments failed to enter the gel (Fig. 3.5a). Even under extended electrophoresis in 0.7% agarose gels using 0.25% TBE as the running buffer no complex entered the gel (results not shown). Furthermore the column purified extract failed to discriminate between ALU148 and the negative control ALU140. Presumably these problems occurred as a result of aggregation of the purified protein. Dilutions of the column purified fusion protein failed to rectify these problems (results not shown). Therefore the purified protein was not used for gel retardation or footprinting experiments.

3.2.2.5 Is the binding of CREA dependent on Zn^{2+} ions?

To determine whether DNA-binding activity is dependent on the presence of Zn^{2+} ions researchers have used the zinc chelator 1,10-phenanthroline to remove available Zn^{2+} ions (Miller *et al.*, 1985; Eisen *et al.*, 1988). In this study 1,10-phenanthroline was incubated with the protein extract for 20 min prior to the addition of the labelled DNA fragment ALU148. Results show (Fig. 3.5b) that 1,10-phenanthroline at a concentration as low as 2 mM was sufficient to abolish binding by GST-CREA. Furthermore, incubation of protein extracts with an excess of Zn^{2+} ions and 600 mM β-mercaptoethanol restored at least partial binding by CREA. Without the β-mercaptoethanol, Zn^{2+} ions did not restore GST-CREA fusion protein binding (results not shown).

3.2.3 Mobility shift analysis of the PST186 region

The *Pst* I-*Pst* I region from -401 to -215, relative to the start point of translation, was analysed for binding to CREA. The plasmid pPST186 was digested with *Hind* III and *Eco* RI and the 186 bp fragment was used in a gel mobility shift assay. The mobility of this fragment using PAGE was retarded in the presence of the GST-CREA fusion protein (Fig. 3.3a). The retardation seen for PST186 was not as complete as that seen for the BP343 fragment, as not all the labelled fragment was retarded, despite the presence of a large excess of fusion protein (section 3.2.2.2).

Further analysis of the PST186 fragment was undertaken by digesting pPST186 with *Dra* I and again with *Hind* III and *Bam* HI. Here the choice of restriction endonucleases was made in order to minimise the amount of polylinker region either side of the insert as well as to allow ease of labelling using the Klenow fragment from DNA polymerase I. Two bands DRA127 and DRA59 were produced and were used in a gel mobility shift assay. The results presented in Figure 3.6a, show that a proportion of the labelled DRA127 fragments formed a retarded complex, whereas the smaller DRA59 fragment showed no retarded mobility. It can be concluded that digesting PST186 with *Dra* I reduced the amount of retardation. When the PST186 fragment was used in several DNase I sensitivity assays no evidence of protection was seen, suggesting that the methods used here were not sufficiently sensitive.

The *Dra* I-*Sma* I fragment (159 bp) from pBAMSAL1.8 was also used in a gel mobility shift assay as it spanned a 40 bp gap (see Fig. 3.2). This fragment displayed no altered mobility in the presence of the CREA fusion protein (Fig. 3.6a).

3.2.4 Mobility shift analysis of the 1.2 kb *Sal* I-*Pst* I region

The *Sal* I-*Pst* I region from approximately -1800 to -401, relative to the start point of translation, was analysed for binding to CREA. Two fragments, *Sph* I- *Acc* I and *Acc* I-*Pst* I(*Sph* I) were obtained by digestion of pPSP1.2 with *Acc* I and *Sph* I . The *Sal* I-*Sph* I fragment was obtained by digesting pPSP1.2 with *Eco* RI and *Sph* I. These DNA fragments were used in gel mobility shift assays with the GST-CREA fusion protein and the results are shown in Figure 3.6b. The (*Eco* RI)*Sal* I-*Sph* I fragment of approximately 660 bp from pPSP1.2 was found to exhibit altered mobility (retardation) in the presence of the CREA fusion protein. The (*Eco* RI)*Sal* I-*Sph* I fragment was sub-cloned and as the sequence for this region had not been determined the plasmid pAMDS660 was sequenced. The sequence data is presented in Figure 3.7 and will be discussed in section 3.9.2. Neither the *Sph* I-*Acc* I fragment nor the *Acc* I-*Pst* I(*Sph* I) fragment showed altered mobility when incubated with the CREA fusion protein.

Figure 3.5a Mobility shift analysis using column purified GST-CREA fusion protein

A mobility shift assay of radiolabelled ALU148 fragment (**Tracks 1-6**) and ALU140 (**Tracks 7-9**) incubated with PBS (**Tracks 1, 4 & 7**), or 150 μg total protein extract containing GST (**Track 2**), or 80 μg total protein extract containing GST-CREA (**Track 3**), or 15 μg of glutathione-sepharose column purified protein extract containing GST (**Tracks 5 & 8**) or 25 μg glutathione-sepharose column purified protein extract containing GST-CREA (**Tracks 6 & 9**).

Figure 3.5b Mobility shift analysis using the fusion protein treated with 1,10-phenanthroline

A mobility shift assay of labelled ALU148 fragment, 5 ng per Track incubated with:

Track 1 PBS

Track 2 300 μg total protein extract containing GST

Track 3 & 4 160 μg total protein extract containing GST-CREA

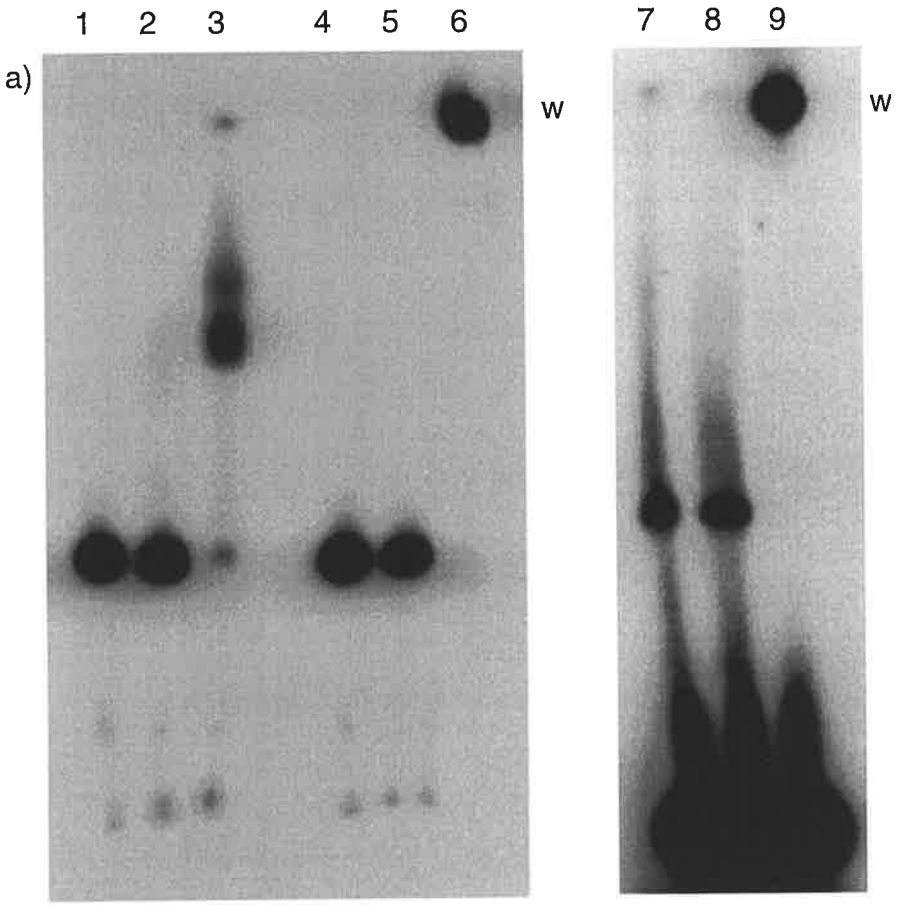
Track 5 160 μg total protein extract containing GST-CREA preincubated with 2 mM 1,10-phenanthroline

Track 6 160 μg total protein extract containing GST-CREA preincubated with 5 mM 1,10-phenanthroline

Track 7 160 μg total protein extract containing GST-CREA preincubated with 10 mM 1,10-phenanthroline

Track 8 160 μg total protein extract containing GST-CREA preincubated with 2 mM 1,10-phenanthroline then with 1M $ZnSO_4$, 600 mM β-mercaptoethanol for 20 min.

Track 9 160 μg total protein extract containing GST-CREA preincubated with 5 mM 1,10-phenanthroline then with 1M $ZnSO_4$, 600 mM β-mercaptoethanol for 20 min.


a)
1 2 3 4 5 6
w
7 8 9
w

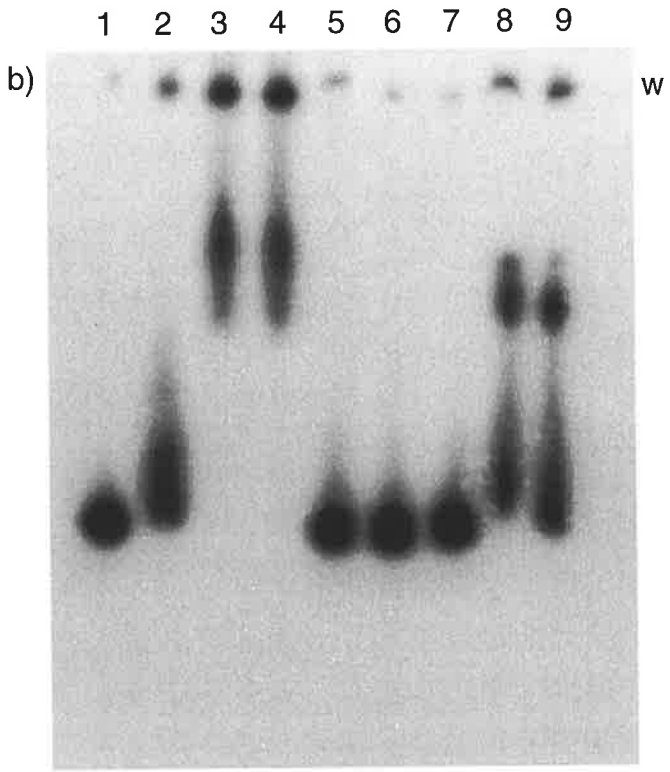

b)
1 2 3 4 5 6 7 8 9
w

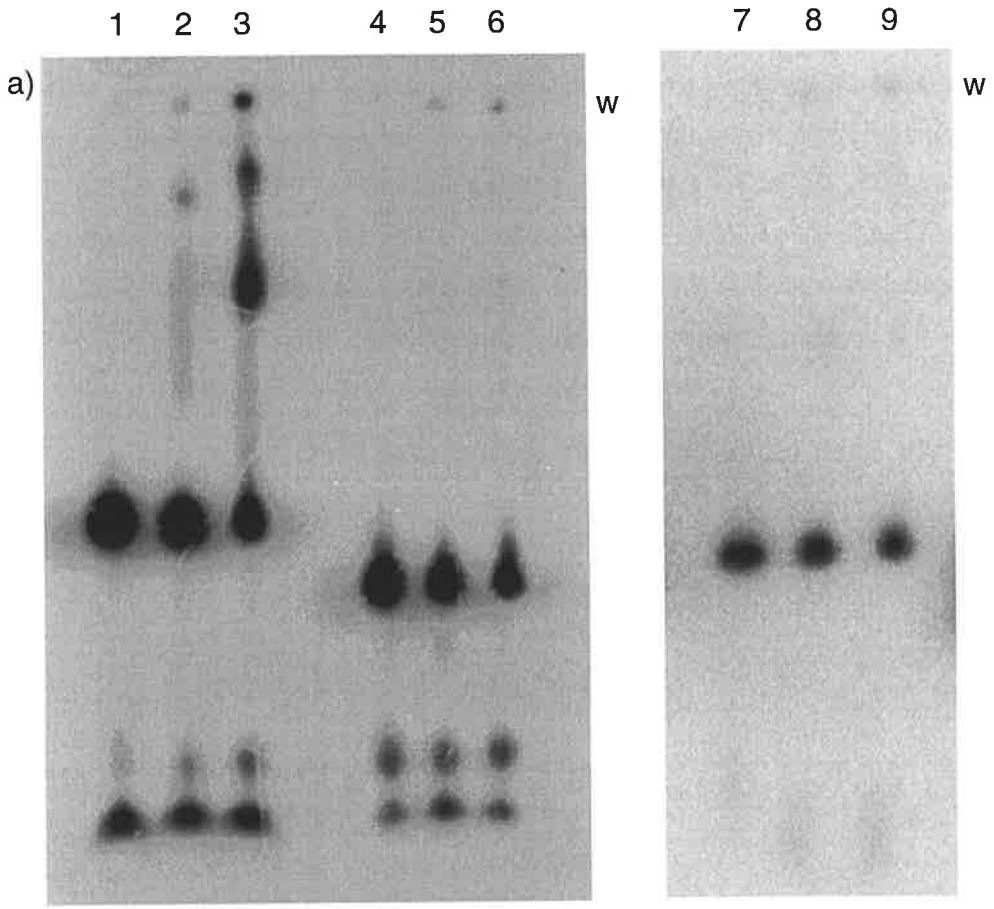
Figure 3.6a Mobility shift analysis of the PST186 region

A mobility shift assay of DRA127 (**Tracks 1-3**), DRA59 (**Tracks 4-6**) and DRA-SMA159 (**Tracks 7-9**). Labelled DNA fragments were incubated with PBS (**Tracks 1, 4 & 7**) or 300 μg total protein extract containing GST (**Tracks 2, 5 & 8**) or 160 μg total protein extract containing GST-CREA (**Tracks 3, 6 & 9**).

Figure 3.6b Further mobility shift analysis on the *amdS* promoter

Fragments are: BP343 (plus 50 bp of polylinker) (**Tracks 1-3**), *Acc* I-*Pst* I (**Tracks 4-6**), *(Eco* RI)*Sal* I-*Sph* I (**Tracks 7-9**) and *Sph* I-*Acc* I (exact size unknown) (**Tracks 10-12**).

Labelled DNA fragments were incubated with PBS (**Tracks 1, 4, 7 & 10**) or 300 μg total protein extract containing GST (**Tracks 2, 5, 8 & 11**) or 160 μg total protein extract containing GST-CREA (**Tracks 3, 6, 9 & 12**).


1 2 3 4 5 6 7 8 9
a)
w
w

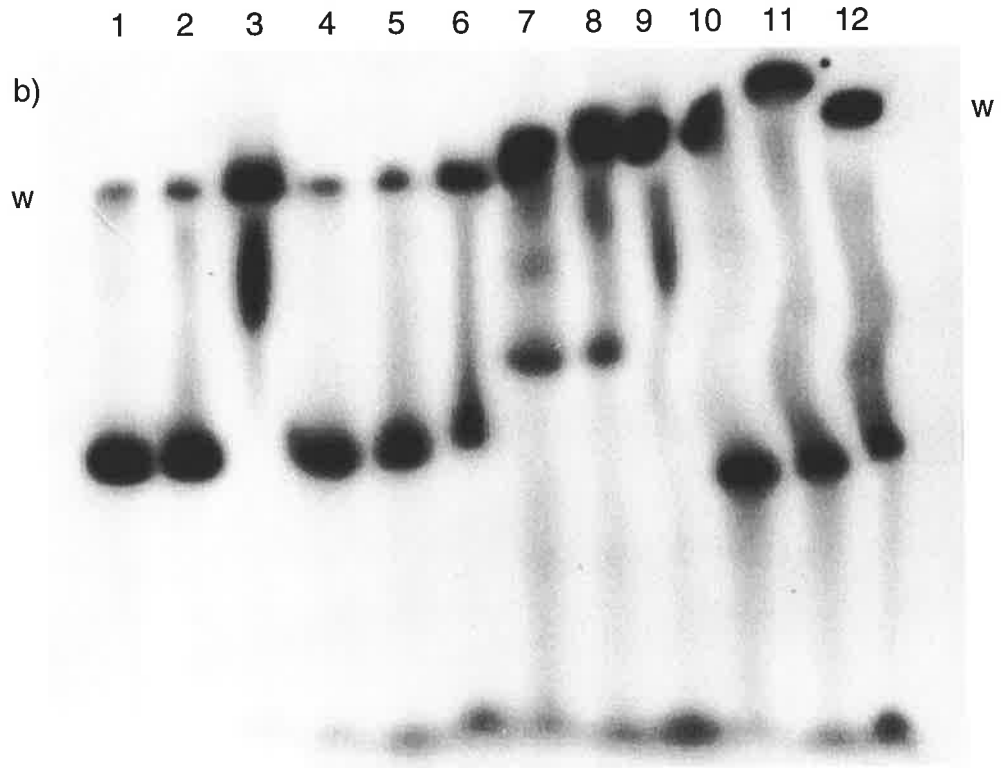
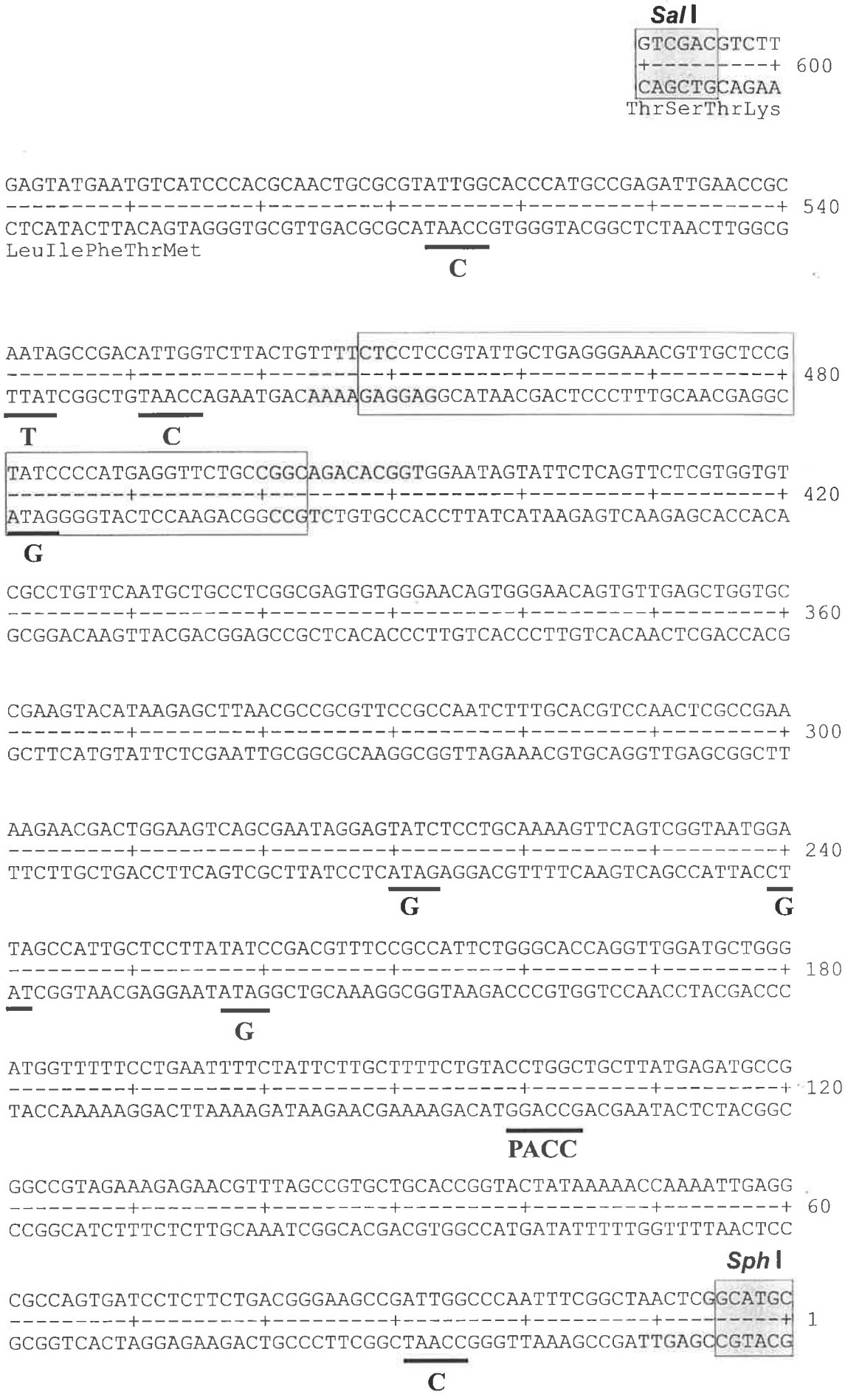

1 2 3 4 5 6 7 8 9 10 11 12
b)
w
w

Figure 3.7 Sequence of the *Sal* I-*Sph* I region from the 5' region of *amdS*

The nucleotide sequence of the 600 bp *Sal* I-*Sph* I region from the 5' region of of *amdS*. Sites for restriction endonucleases are shown above the sequence and numbering of the nucleotides is shown on the right. The open reading frame and the position of CCAAT (C; van Heeswijck and Hynes, 1991), GATA/ TATC (G; Merika and Orkin, 1993), TATA-like (T; Gurr *et al.*, 1987) and GCCA G/A G (PACC; Tilburn *et al.*, 1995) sequences (solid lines) are shown for a putative gene transcribed in the opposite direction to *amdS*. A region discussed in section 3.9.2 is boxed. The sequence was determined in one direction only.


SalI
GTCGACGTCTT
CAGCTGCAGAA
ThrSerThrLys
600
GAGTATGAATGTCATCCCACGCAACTGCGCGTATTGGCACCCATGCCGAGATTGAACCGC
CTCATACTTACAGTAGGGTGCGTTGACGCGCATAACCGTGGGTACGGCTCTAACTTGGCG
LeuIlePheThrMet
C
540
AATAGCCGACATTGGTCTTACTGTTTTCTCCTCCGTATTGCTGAGGGAAACGTTGCTCCG
TTATCGGCTGTAACCAGAATGACAAAAGAGGAGGCATAACGACTCCCTTTGCAACGAGGC
T
C
480
TATCCCCATGAGGTTCTGCCGGCAGACACGGTGGAATAGTATTCTCAGTTCTCGTGGTGT
ATAGGGGTACTCCAAGACGGCCGTCTGTGCCACCTTATCATAAGAGTCAAGAGCACCACA
G
420
CGCCTGTTCAATGCTGCCTCGGCGAGTGTGGGAACAGTGGGAACAGTGTTGAGCTGGTGC
GCGGACAAGTTACGACGGAGCCGCTCACACCCTTGTCACCCTTGTCACAACTCGACCACG
360
CGAAGTACATAAGAGCTTAACGCCGCGTTCCGCCAATCTTTGCACGTCCAACTCGCCGAA
GCTTCATGTATTCTCGAATTGCGGCGCAAGGCGGTTAGAAACGTGCAGGTTGAGCGGCTT
300
AAGAACGACTGGAAGTCAGCGAATAGGAGTATCTCCTGCAAAAGTTCAGTCGGTAATGGA
TTCTTGCTGACCTTCAGTCGCTTATCCTCATAGAGGACGTTTTCAAGTCAGCCATTACCT
G
G
240
TAGCCATTGCTCCTTATATCCGACGTTTCCGCCATTCTGGGCACCAGGTTGGATGCTGGG
ATCGGTAACGAGGAATATAGGCTGCAAAGGCGGTAAGACCCGTGGTCCAACCTACGACCC
G
180
ATGGTTTTTCCTGAATTTTCTATTCTTGCTTTTCTGTACCTGGCTGCTTATGAGATGCCG
TACCAAAAAGGACTTAAAAGATAAGAACGAAAAGACATGGACCGACGAATACTCTACGGC
PACC
120
GGCCGTAGAAAGAGAACGTTTAGCCGTGCTGCACCGGTACTATAAAAACCAAAATTGAGG
CCGGCATCTTTCTCTTGCAAATCGGCACGACGTGGCCATGATATTTTTGGTTTTAACTCC
60
SphI
CGCCAGTGATCCTCTTCTGACGGGAAGCCGATTGGCCCAATTTCGGCTAACTCGGCATGC
GCGGTCACTAGGAGAAGACTGCCCTTCGGCTAACCGGGTTAAAGCCGATTGAGCCGTACG
C
1

3.2.5 Summary

From these results the binding of the CREA fusion protein to the 5' region of *amdS* was localised to three regions spanned by the ALU148 fragment, the DRA127 fragment and the 660 bp (*Eco* RI)*Sal* I-*Sph* I fragment much further upstream. The GST-CREA fusion protein did not bind DRA127 with as much affinity as it did the larger fragment PST186. Further analyses of binding to *amdS* focused on the ALU148 fragment due to its complete retardation and proximity to the start point of transcription.

3.3 DNase I Sensitivity analysis of the ALU148 fragment from *amdS*

In order to carry out DNase I sensitivity assays (footprinting) the ALU148 fragment was cloned into the *Sma* I site of pUC19 (pALU148). For analysis of the coding strand, pALU148 was digested with *Bam* HI followed by end labelling and digestion with *Kpn* I. The *Bam* HI-*Kpn* I fragment was then purified by non-denaturing PAGE and used for footprinting analysis, as described in section 2.2.6. The two electrophoretically distinct complexes formed with CREA protein (HMC and LMC) were excised separately and electrophoresed in separate tracks in the final analytical gel. The result is shown in Figure 3.8. There was endogenous exonuclease activity in the *E. coli* GST-CREA total protein extract under the conditions used for DNase I digestion (5 mM Ca^{2+}, 10 mM Mg^{2+}). Consequently radiolabel was lost from many DNA molecules. Endfilling of the *Bam* HI site with one labelled adenine and subsequent attack by exonuclease results in fragments which migrate one base shorter than the size of those in the Maxam and Gilbert sequencing and the control tracks (Fig. 3.8). The HMC has a similar pattern of degradation to the fragment incubated without CREA except that two "windows" of protection are present. The 20 bp and 17 bp windows correspond to the protection of bases from -81 to -100 and from -106 to -121 respectively. The protected regions are shown above the sequence in Figure 3.10. There is one base within and many outside the protected region which show increased sensitivity to DNase I digestion (ie are hyperdigested). Although a very low level of radioactivity was present in the LMC track, a similar pattern of protection, to the HMC track, was seen.

Binding of the GST-CREA fusion protein to the non-coding strand of ALU148 was analysed by using an *Eco* RI-*Bam* HI fragment (Fig. 3.9a) and an *Eco* RI-*Nla* III sub-fragment (Fig. 3.9b). The larger *Eco* RI-*Bam* HI fragment was analysed by digesting pALU148 with *Eco* RI and *Hind* III and end labelling. Further digestion with *Bam* HI produced an *Eco* RI-*Bam* HI fragment labelled only at the *Eco* RI end. After purification, footprinting was carried out, except that the two electrophoretically distinct CREA-DNA complexes were not excised separately. The results show that a similar pattern of degradation was seen in all tracks (Fig. 3.9a) except that a large (62 bp) and small (18 bp) window of protection is visible within the **R** track. The windows indicate that the sequence from -41 to -58 and from -60 to -121 is protected and this is shown by the shaded boxes below the sequence in Figure 3.10. There was no difference in the pattern of protection seen with the lower and higher concentrations of DNase I used.

The smaller *Eco* RI-*Nla* III fragment was analysed by digesting pALU148 with *Eco* RI and *Pst* I and end labelling. Further digestion with *Nla* III produced an *Eco* RI-*Nla* III fragment labelled only at the *Eco* RI end. The two electrophoretically distinct complexes were excised separately and the results are shown in Figure 3.9b. Since end filling of the *Eco* RI "sticky" end results in two labelled adenines being present, partial digestion with contaminating exonuclease results in an additional fragment one base shorter. This is visible as a doublet for each band on the sequencing gel and makes interpretation of band intensity at each position complicated. The pattern of degradation for this fragment in the presence of CREA is very different to that in the absence of CREA. This was not seen for previous footprinting experiments and makes it difficult to locate windows accurately. The pattern of degradation for the low and high mobility complexes is somewhat similar. There is one large window of protection from -119 to -56 (the end of the fragment) and within this region there are at least seven hyperdigested bases. The two concentrations of DNase I give similar but not identical results. For example, at the lower concentration of DNase I a window extends from -134 to -158+polylinker but at a higher concentration only the -146 to -158+polylinker region appears to be protected. This protection is in the form of two separate windows (-146 to -151 and -155 to -158+polylinker) and the three

Figure 3.8 DNase I sensitivity assay (footprinting) on the coding strand of the ALU148 fragment

The ALU148 fragment from *amdS*, labelled at the *Bam* HI end, was incubated with 100 ng of DNase I in the presence (**H** and **L**) or absence (**F**) of the GST-CREA fusion protein and electrophoresed on a mobility shift gel. Bands corresponding to the free fragment (**F**), the higher mobiliy complex (HMC) (**H**) and the lower mobility complex (LMC) (**L**) were excised and eluted. After purification samples were electrophoresed on a sequencing gel alongside a Maxam and Gilbert "G" sequencing reaction (**S**).

The sequence of the fragment pictured is shown on the left and alongside it the two protected regions of the **H** complex are shown as open boxes. Hyperdigested bases are marked with an *, unprotected bases are shown with a straight line and a summary of these findings is shown in Figure 3.10

DNA fragments in the **H** and **L** tracks migrate one base shorter due to endogenous nuclease acitvity.


CTAAACCCTATAATTAG
TCTCTTATCAACACCATCCGCTCCCCCGGGATCAATGAGGAGAATGAGGGGATGCGGGGCTAAAGAAGCCTACATAACCCTCATGCCAACTCCCAGTTTACACTCGTCGAGCCA

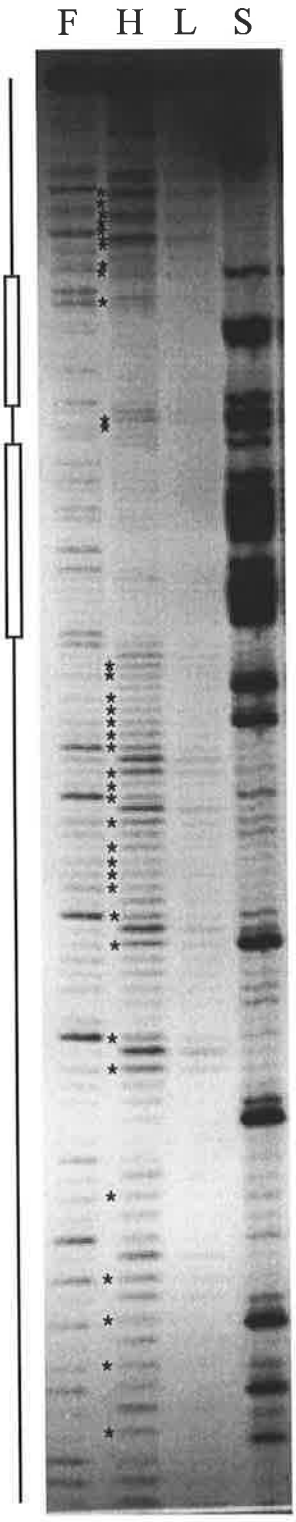

F H L S

Figure 3.9a/ b DNase I sensitivity assays (footprinting) on the non-coding strand of the ALU148 fragment

The ALU148 fragment from *amdS* (**a**) labelled at the *Eco* RI end was incubated with 50 ng or 100 ng of DNase I in the presence (**R**) or absence (**F**) of the GST-CREA fusion protein and electrophoresed on a mobility shift gel. Bands corresponding to the free fragment (**F**) and retarded complexes (**R**) were excised and eluted. After purification samples were electrophoresed on a sequencing gel alongside a Maxam and Gilbert "G" sequencing reaction (**S**).

The smaller *Eco* RI - *Nla* III sub-fragment of ALU148 (**b**) labelled at the *Eco* RI end was incubated with 50 ng or 100 ng of DNase I in the presence (**H** and **L**) or absence (**F**) of the GST-CREA fusion protein and electrophoresed on a mobility shift gel. Bands corresponding to the free fragment (**F**), the higher mobiliy complex (HMC) (**H**) and the lower mobility complex (LMC) (**L**) were excised and eluted. After purification samples were electrophoresed on a sequencing gel alongside a Maxam and Gilbert "G" sequencing reaction (**S**).

The sequence of the smaller sub-fragment is shown on the right and that of the larger fragment on the left. The protected regions of the **H** complex for the 100 ng DNase I track and the **R** complex are shown as open boxes. The hyperdigested bases are marked with an * next to the sequence and the **H** (100 ng DNase I) track. DNA fragments in the **H** and **L** tracks migrate as doublets due to endogenous nuclease acitvity. A summary of these findings is shown in Figure 3.10.

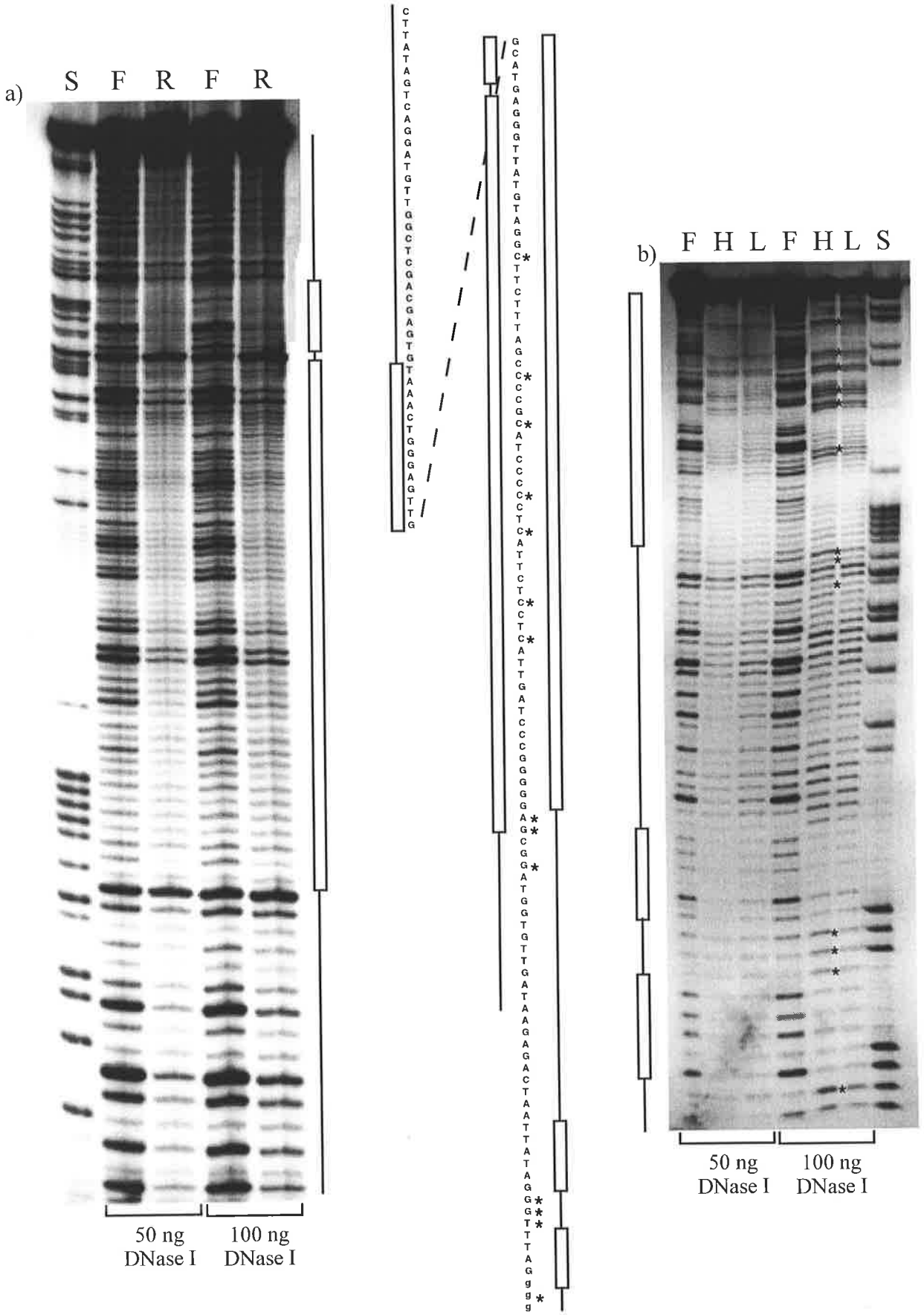
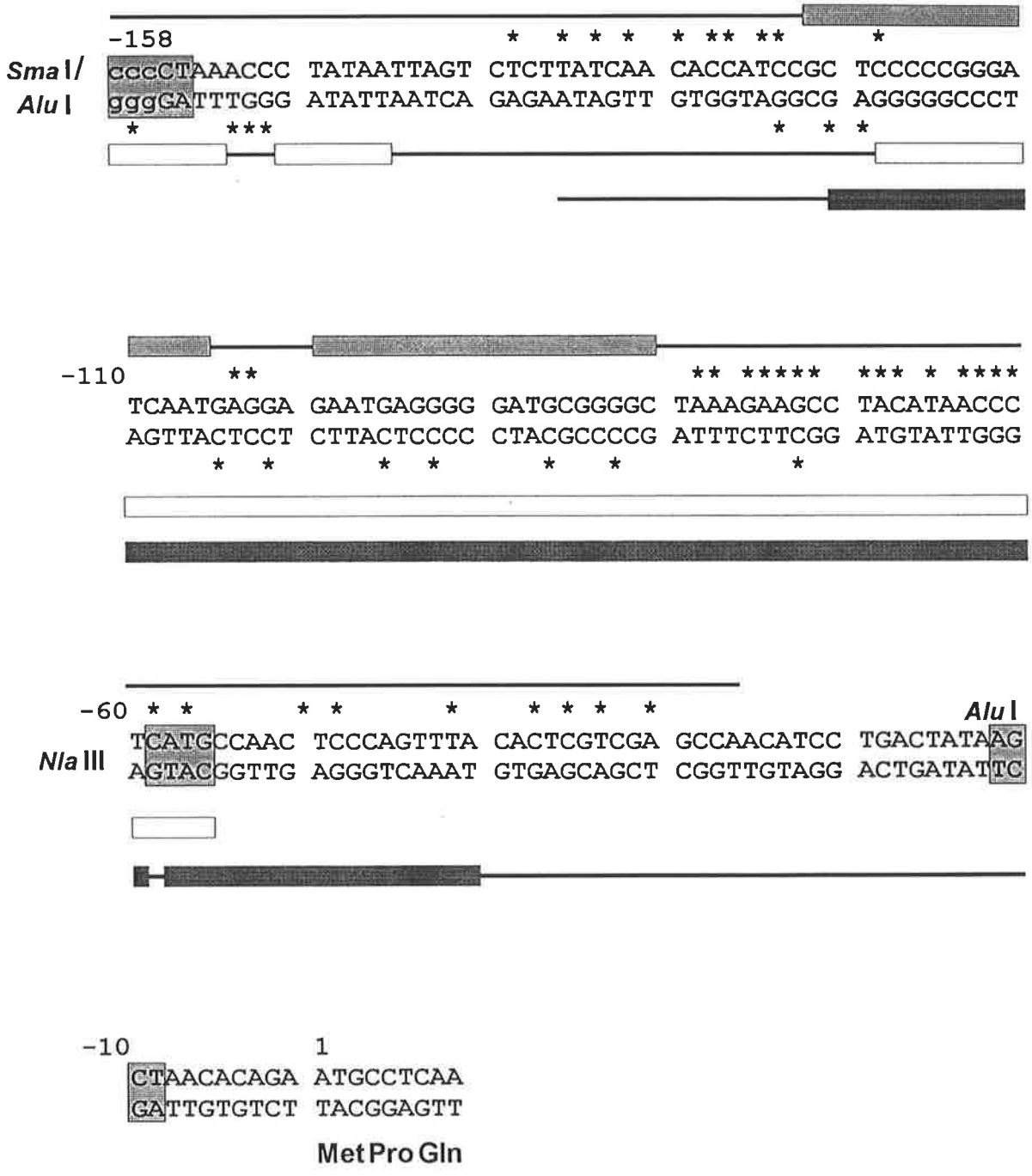

a)
S F R F R
50 ng DNase I
100 ng DNase I
b)
F H L F H L S
50 ng DNase I
100 ng DNase I

Figure 3.10 Summary of the protected regions of the ALU148 fragment

The sequence of the ALU148 fragment from *amdS* including some restriction endonuclease sites and the nucleotide positions relative to the start point of translation are shown. The protected regions on the coding strand are shown above the sequence and the protected regions for the non-coding strand are shown below the sequence. For the non-coding strand lighter boxes represent regions protected in the HMC (100 ng DNase I, Fig. 3.9b) and the darker boxes represent regions protected in the **R** complex (Fig. 3.9a). Hyperdigested bases are marked with an *.


-158
SmaI/
AluI
cccCTAAACCC TATAATTAGT CTCTTATCAA CACCATCCGC TCCCCCGGGA
gggGATTTGGG ATATTAATCA GAGAATAGTT GTGGTAGGCG AGGGGGCCCT
-110
TCAATGAGGA GAATGAGGGG GATGCGGGGC TAAAGAAGCC TACATAACCC
AGTTACTCCT CTTACTCCCC CTACGCCCCG ATTTCTTCGG ATGTATTGGG
-60
AluI
NlaIII
TCATGCCAAC TCCCAGTTTA CACTCGTCGA GCCAACATCC TGACTATAAG
AGTACGGTTG AGGGTCAAAT GTGAGCAGCT CGGTTGTAGG ACTGATATTC
-10
1
CTAACACAGA ATGCCTCAA
GATTGTGTCT TACGGAGTT
Met Pro Gln

bases between them are hyperdigested. The pattern of protection corresponding to the HMC at the higher DNase I concentration is shown as open boxes below the sequence in Figure 3.10. An analysis of the sequence protected by CREA on both strands shows that this region is very GC rich. These findings will be discussed further in section 3.8.2 and 3.9.2.

3.4 Mobility shift analysis of clones containing altered *amdS* sequences

A detailed genetical and molecular analysis of the 5' controlling region of *amdS* has provided a number of *cis*-acting mutations. Three of these were investigated further as they lie within the protected region of ALU148. The region containing the *amdI18* mutation was used in a gel mobility shift assay to determine whether this mutation affected CREA binding. The plasmid p718B16-322 containing the *amdI18* mutation was digested with *Bam* HI and *Pst* I and the 353 bp fragment used in a mobility shift assay. Another mutation was also tested to see if it disrupted CREA binding. The plasmid pLIT14, which contains the *amdI9* mutation and a *Bam* HI linker inserted into the *Sma* I site of pLIT1 (Littlejohn and Hynes, 1992), was digested with *Bam* HI and *Pst* I and the 353 bp fragment was used in a gel mobility assay. The control fragment, BP343, from wild type *amdS* was also tested and the results are shown in Figure 3.11. Both fragments containing mutations within the protected region of ALU148 are able to bind CREA and do so with an affinity similar to the wild type fragment of *amdS*.

3.5 Mobility shift analysis of the upstream sequences of *facB*

Evidence from molecular genetic manipulations has shown that *facB* is under CREA control (discussed in section 1.3.2). An investigation of binding by the CREA fusion protein to the *facB* promoter region was undertaken to determine whether a direct role for the CREA protein in *facB* regulation was possible and to compare binding sites with those found within the *amdS* promoter. A map of the 5' region of *facB* and the restriction sites used in this study, is shown in Figure 3.12. Several fragments from the 5' region of *facB* were sub-cloned to provide sufficient material for analysis and these are also shown. DNA fragments from the 5' region of

facB, starting with those closest to the start point of transcription, were tested for their ability to bind the CREA fusion protein.

3.5.1 Mobility shift analysis of the region 742 bp upstream from the start point of translation of *facB*

The *Nru* I-*Sal* I region -186 to +173, relative to the start point of translation, was assayed for binding by the CREA fusion protein as an *Nru* I-*Sal* I fragment of 359 bp from pFAB5-1 Sac SK+-9HdIII self-4. The 359 bp fragment was digested with *Sma* I to produce two fragments, one 153 bp and the other 206 bp. As shown in Figure 3.13 neither fragment showed evidence of strong binding to CREA. There did appear to be a slight retarded band but given the large excess of protein used, this was not considered significant. To check that *Sma* I digestion had not destroyed part of the CREA binding site, the *Nru* I-*Sal* I fragment was digested with *Eco* RI. Two fragments 194 bp and 165 bp were assayed for binding by the CREA fusion protein (Fig. 3.13). Again no retardation was seen for either of these two fragments.

Further upstream, the *Sac* I-*Nru* I region -569 to -186 was analysed by sub-cloning the *Sac* I-*Nru* I fragment of 383 bp (pNE383). pNE383 was digested with *Eco* RI and *Bam* HI and the 400 bp insert isolated. Further digestion with *Nla* III produced three fragments which were isolated on a 4% metaphor agarose gel. The 184 bp and 146 bp fragments were labelled using the Klenow fragment from DNA polymerase I and the small fragment of 53 bp was labelled using polynucleotide kinase (section 2.2.5.2 procedure (B)). All three fragments were used in gel mobility shift experiments and no significant retardation was seen for the 184 bp and 53 bp fragments (Fig. 3.13). The 146 bp fragment did show some retardation, however, as the pattern of retardation was identical in the GST and GST-CREA containing tracks, this was not considered to be due to CREA. To ensure that a possible site was not destroyed by the restriction digest, the plasmid pNE383 was digested with *Bam* HI and *Eco* RI. The 385 bp fragment was purified using agarose gel electrophoresis and then redigested with *Taq* I restriction endonuclease. The three fragments of 58, 106, and 219 bp were isolated on a 4%

Figure 3.11 Mobility shift analysis of mutant 5' *amdS* sequences

A mobility shift assay of the *Bam* HI-*Pst* I fragments from the *in vitro* generated mutation pLIT14 (**Tracks 1-3**), from wild type (**Tracks 4-6**), and from the *amdI18* strain (**Tracks 7-9**) of *A. nidulans*.

Labelled DNA fragments were incubated with PBS (**Tracks 1, 4 & 7**) or 300 μg total protein extract containing GST (**Tracks 2, 5 & 8**) or 160 μg total protein extract containing GST-CREA (**Tracks 3, 6 & 9**).

The same protein extracts were used for all three fragments, yet the GST containing extract produced significant retardation only with the BP353 (wild type) fragment. As the pGEX-2T extract used had not been thawed, binding to the BP353 fragment was expected as discussed earlier. Why there was no binding to the *amdI18* or pLIT14 derived fragments is not clear.

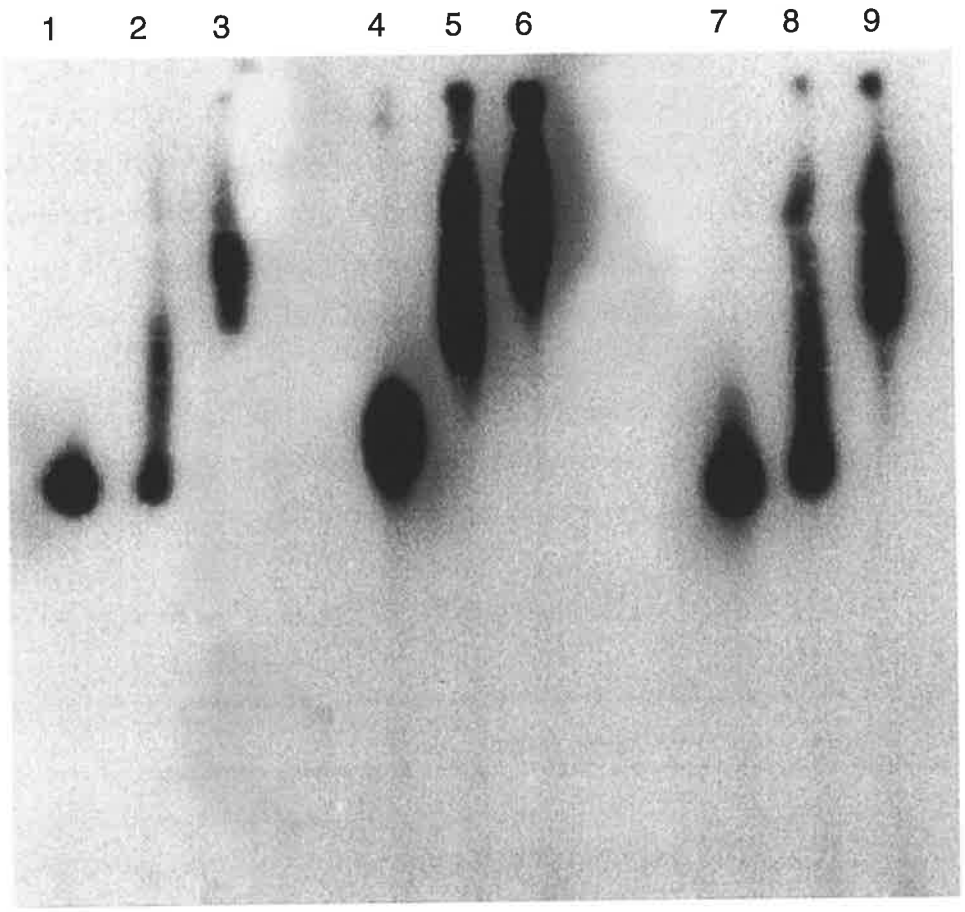

1
2
3
4
5
6
7
8
9

Figure 3.12 Partial restriction endonuclease map of the 5' region of *facB*

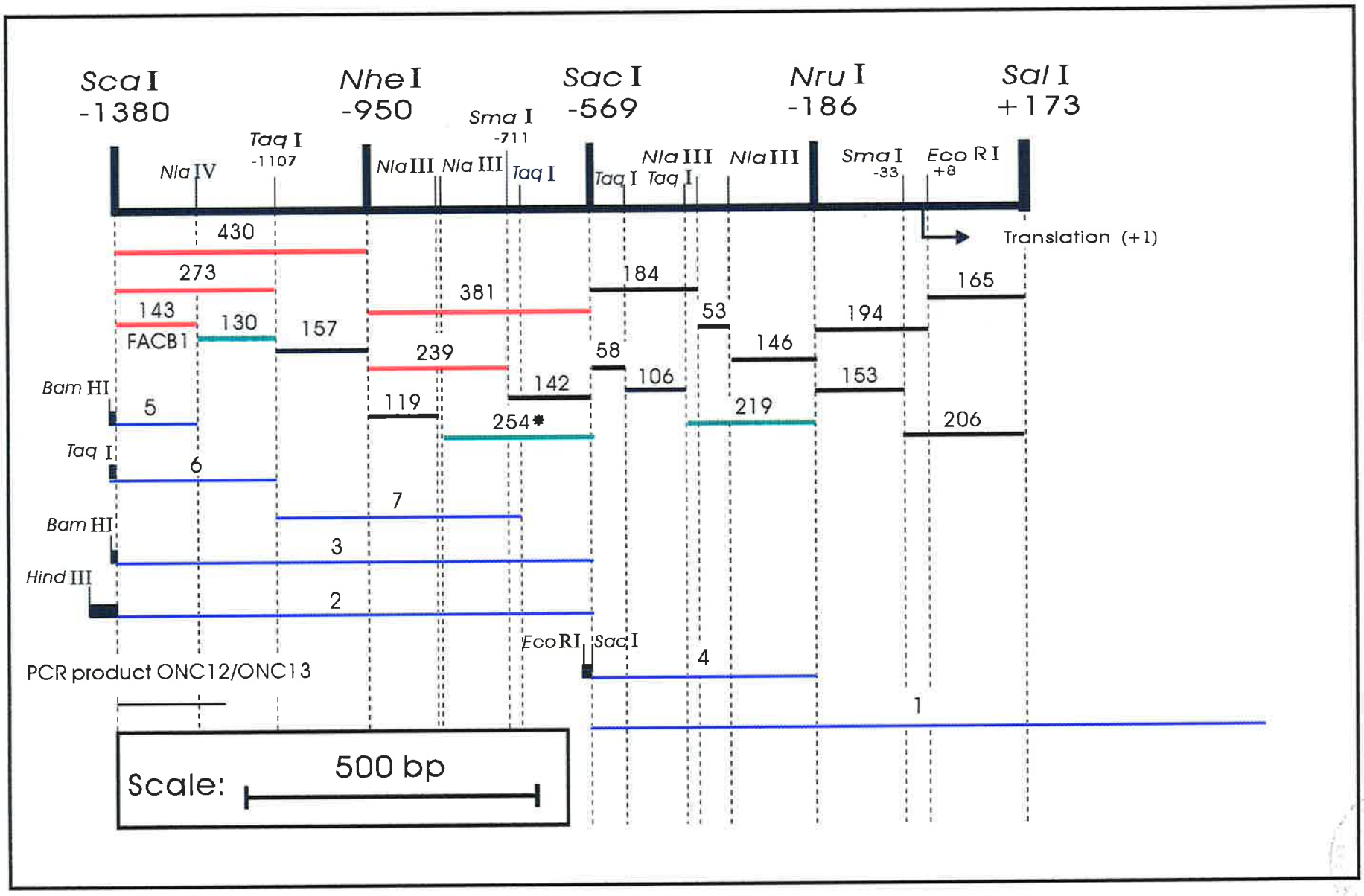
Scale map of the 5' region of the *facB* gene (Todd *et al.*, submitted). Fragments isolated for mobility shift assays are shown as horizontal bars with sizes (excluding polylinker regions from sub-clones) marked above the bars. The start point of transcription is not known (M.J. Hynes pers. comm.). Fragments which were found to bind GST-CREA strongly are shown in red. Fragments that did not bind strongly but which may bind weakly are shown in green. For one of these (marked with an *) binding to GST-CREA could not be assesed due to strong binding by the GST control extract. The start point of translation is +1.

Blue horizontal lines indicate the clones constructed from the 5' region of the *facB* gene. Numbers above these lines refer to the Table below. The heavy lines represent regions of polyliker with restriction endonuclease sites used in the cloning procedure.

Plasmid clones constructed from the 5' region of the *facB* gene

Number	Name	Cloning Description
1	p1.2FACB	The 1.2kb *Sac* I-*Hind* III fragment from pFAB5-1 Sac SK+-9HIII self-4 cloned into pUC19 digested with *Sma* I (Sequencing revealed *Hind* III site and 10bp before it is missing)
2	p800PUBS	The *Hind* III-*Sac* I fragment of approx. 800 bp from pFAB5-1 was cloned into the *Sma* I site of pUBS*
3	p800PUBS-BAM	The *Bam* HI-*Bam* HI polylinker region of p800PUBS was removed and the plasmid religated
4	pNE383	The *Eco* RI-*Nru* I fragment from pFAB5-1 Sac SK+-9HIII self-4 was cloned into pUC19 digested with *Eco* RI and *Sma* I
5	pFACB1	The *Bam* HI-*Nla* IV fragment of p800PUBS was cloned into pUC19 digested with *Bam* HI and *Sma* I
6	pFACB2	285 bp *Taq* I(*Sca* I)-*Taq* I (-1107) from pFAB5-1 cloned into pUBS* digested with *Eco* RV
7	pFACB3	421 bp *Taq* I(-1107)-*Taq* I (-685) from pFAB5-1 cloned into pUBS* digested with *Eco* RV

*The pUBS vector contains the pBLUESCRIPT multiple cloning region in place of the pUC19 multiple cloning region.


Sca I
-1380
Nhe I
-950
Sac I
-569
Nru I
-186
Sal I
+173
Sma I
-711
Taq I
-1107
Nla IV
Nla III
Nla III
Taq I
Taq I
Taq I
Nla III
Nla III
Sma I
-33
Eco R I
+8
Translation (+1)
430
273
143
FACB1
130
157
381
239
142
119
254*
184
165
53
194
146
58
106
153
219
206
Bam HI
5
Taq I
6
7
Bam HI
3
Hind III
2
Eco RI
Sac I
4
1
PCR product ONC12/ONC13
Scale:
500 bp

metaphor agarose gel. The 106 bp fragment could not be labelled with radiolabelled dATP but substituting radiolabelled dCTP permitted labelling with the Klenow fragment from DNA polymerase I. As can be seen from Figure 3.13, fragments 58 and 106 did not retard. Interpretation for the 219 bp fragment is complicated by the binding seen to protein(s) in the GST containing extract and therefore some retardation can not be ruled out.

3.5.2 Mobility shift analysis of fragments further 5' of *facB*

Since a clearly retarding fragment from the 5' region of *facB* was not detected within 700 bp of the start point of translation a larger clone of the 5' region was analysed. A fragment of the 5' region of *facB* from the *Sca* I to the *Sac* I site of pFAB5-1 (approximately 800 bp) was sub-cloned into the *Sma* I site of pUBS. The resulting plasmid was called p800PUBS. The *Nhe* I-*Sac* I fragment (from -950 to -569 5' of *facB*) was excised as a *Nhe* I-*Bam* HI digested product and purified by agarose gel electrophoresis. This fragment was labelled and showed strong retardation. It was digested further with *Sma* I producing two fragments of 239 and 142 bp. In gel retardation assays the larger fragment *Nhe* I-*Sma* I gave good retardation but the smaller *Sma* I-*Sac* I(*Bam* HI) fragment exhibited no retardation (Fig. 3.14). The *Nhe* I-*Sac* I region was checked by digestion with *Nla* III to give two fragments one 119 bp and the other 254 bp (missing an 8 bp fragment between the two *Nla* III sites). As shown in Fig. 3.14 the smaller fragment did not retard and the larger one showed binding by pGEX-2T control extracts. This binding did not reduce upon freezing and thawing and thus binding to GST-CREA could not be assessed but is likely given the binding of the 239 bp fragment.

The *Sca* I-*Nhe* I region from -1380 to -950 was tested for binding by CREA. p800PUBS-Bam (see Fig. 3.12) was digested with *Eco* RI and *Nhe* I and the 430 bp band was purified. Further digestion of this band with *Taq* I gave two fragments which were labelled using the Klenow fragment of DNA polymerase I. These were assayed for binding to GST-CREA (Fig 3.14). The *Taq* I-*Nhe* I band of 157 bp showed no retarded mobility whereas the *Sca* I-*Taq* I band of 273 bp showed strong retardation. The 273 bp region was divided into two sub-fragments by

Figure 3.13 Mobility shift analysis of the fragments within the first 742 bp 5' of *facB*

Mobility shift assay of fragments from the 5' *facB* region:

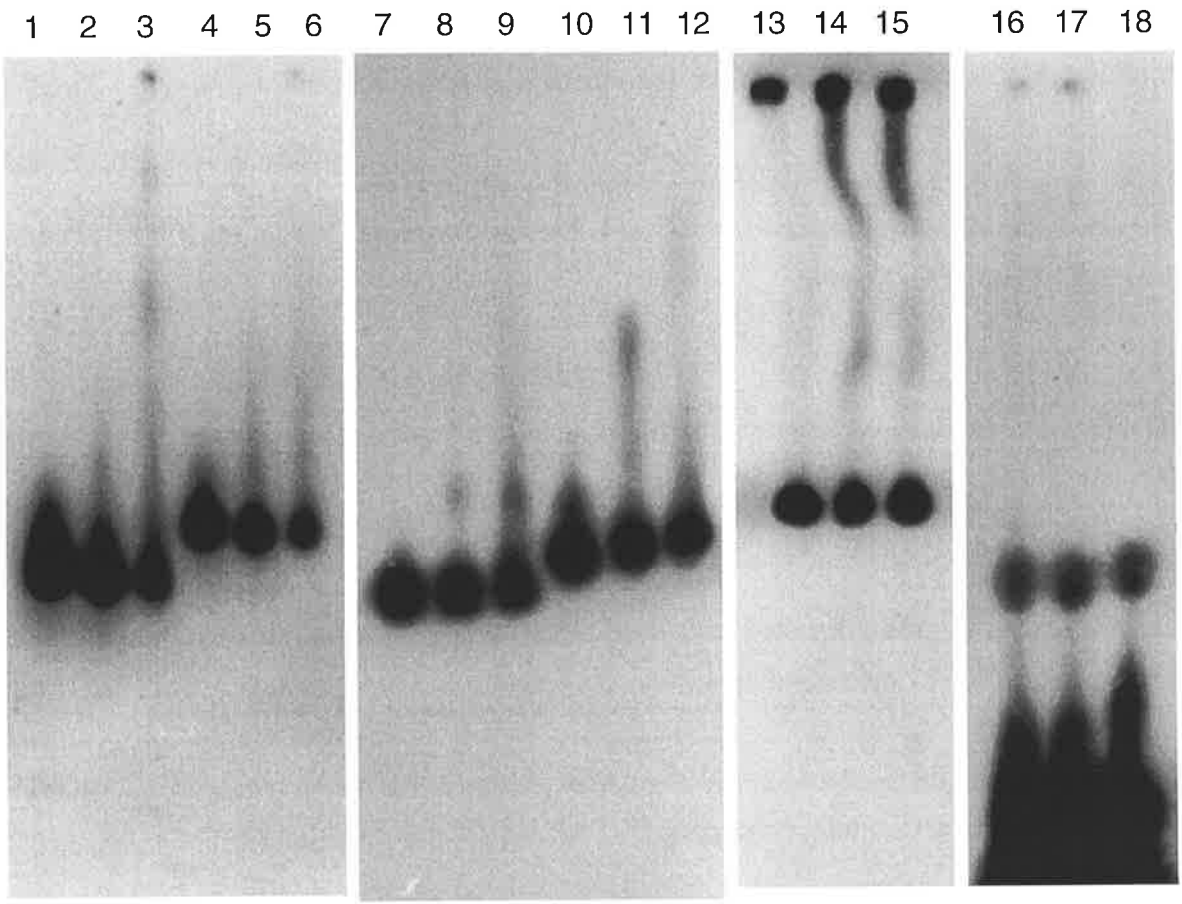
The *Nru* I -*Sal* I fragment digested with *Sma* I.
The two fragments shown are 153 bp (**Tracks 1-3**) and 206 bp (**Tracks 4-6**).

The *Nru* I-*Sal* I fragment digested with *Eco* RI.
The two fragments shown are 165 bp fragment (**Tracks 7-9**) and 194 bp (**Tracks 10-12**).

The *Sac* I-*Nru* I fragment digested with *Nla* III.
The three fragments shown are 184 bp (**Tracks 13-15**), 53 bp (**Tracks 16-18**) and 146 bp (**Tracks 19-21**).

The *Sac* I-*Nru* I fragment digested with *Taq* I. The three fragments shown are 58 bp (**Tracks 22-24**), a longer exposure of 106 bp (overlayed) (**Tracks 25-27**) and 219 bp (**Tracks 28-30**).

In all cases the first Track contains labelled fragment incubated with PBS, the second with 300 μg total protein extract containing GST and the third Track with 160 μg total protein extract containing GST-CREA.


1 2 3 4 5 6
7 8 9 10 11 12
13 14 15
16 17 18

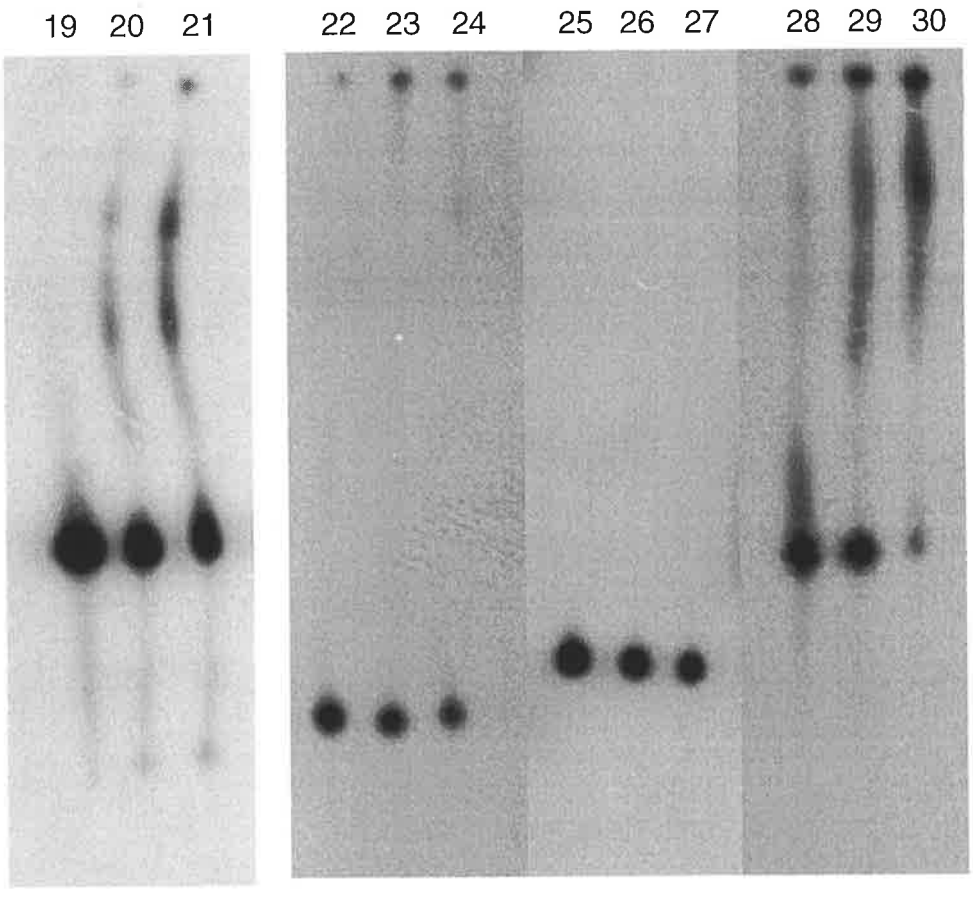

19 20 21
22 23 24
25 26 27
28 29 30

Figure 3.14 Mobility shift analysis on fragments further upstream of *facB*

Mobility shift assay of fragments from the 5' *facB* region (see also Fig. 3.12):

The *Nhe* I-*Sac* I fragment digested with *Sma* I. The two fragments shown are 239 bp (**Tracks 1-3**) and 142 bp (**Tracks 4-6**).

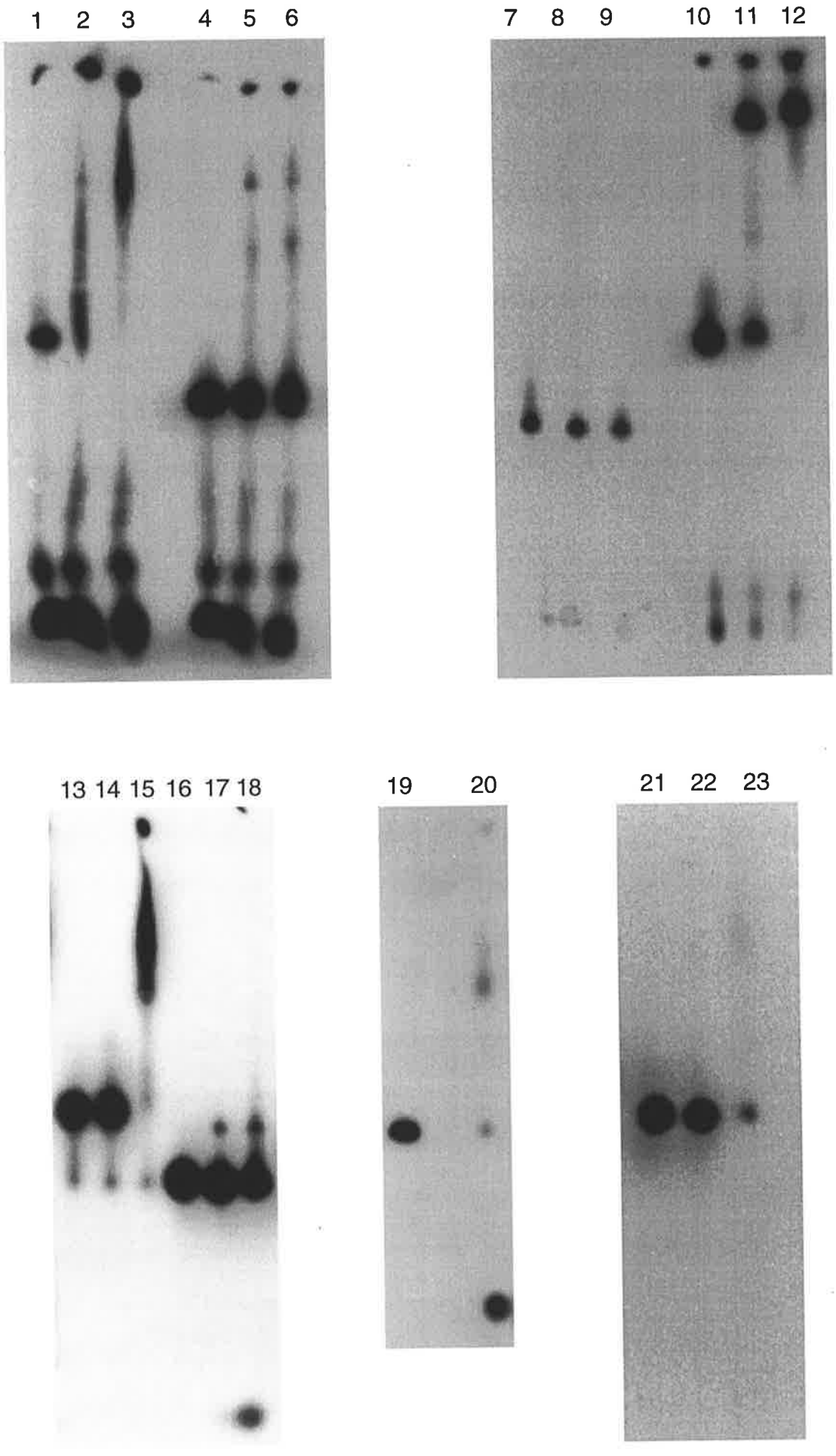
The *Nhe* I-*Sac* I I fragment digested with *Nla* III. The two fragments shown are 119 bp (**Tracks 7-9**) and 254 bp (**Tracks 10-12**).

The *Sca* I-*Nhe* I fragment digested with *Taq* I. The two fragments shown are 273 bp (**Tracks 13-15**), and 157 bp (**Tracks 16-18**).

The FACB1 fragment incubated with 300 µg total protein extract containing GST (**Track 19**) or 160 µg total protein extract containing GST-CREA (**Track 20**).

The *Nla* IV-*Taq* I fragment of 130 bp (**Tracks 21-23**).

In all cases except for FACB1 the first Track contains labelled fragment incubated with PBS, the second with 300 µg total protein extract containing GST and the third Track with 160 µg total protein extract containing GST-CREA.


1 2 3 4 5 6
7 8 9 10 11 12
13 14 15 16 17 18
19 20
21 22 23

digestion with *Nla* IV. The 143 bp FACB1 fragment was sub-cloned into pUC19 (pFACB1) and tested for binding to CREA. In this particular binding reaction 5 mM Ca^{2+} and 10 mM Mg^{2+} was present and therefore attack by endogenous nuclease was significant, however FACB1 showed good retardation (Fig. 3.14) . The 130 bp *Nla* IV- *Taq* I fragment showed very slight retardation with the CREA fusion protein.

3.6 DNase I sensitivity assays of the FACB1 fragment

The region that produced the strongest binding to CREA was the 273 bp fragment. This fragment was too large for use in footprinting, therefore the FACB1 fragment which showed good retardation was used. Footprinting was carried out to determine which bases within FACB1 were protected by the CREA fusion protein.

For the non-coding strand p800PUBS-Bam was digested with *Bam* HI, labelled and digested with *Nhe* I and *Nla* III. The *Bam* HI-*Nla* III fragment labelled at the *Bam* HI end, was purified and footprinted. The resulting pattern of degradation (Fig. 3.15a) is similar in all four tracks, except that two windows of protection, 25 bp and 18 bp in length, are visible for the lower concentration of DNase I in the **R** track. One of these is a region of strong protection from -1345 to -1369 and the other is a region of weaker protection from -1300 to -1317. These results are shown below the sequence in Figure 3.16.

The coding strand was footprinted by digesting pFACB1 with *Hind* III and *Eco* RI. This fragment was purified by agarose gel electrophoresis and labelled using the Klenow fragment of DNA polymerase I. After further digestion with *Bam* HI and purification, the *Eco* RI-*Bam* HI fragment labelled at the *Eco* RI end was footprinted (Fig. 3.15b). The pattern of degradation of tracks with and without CREA is similar except where two windows of strong protection, and a third of weaker protection, are present. One region of strong protection extends from -1348 to -1365. The other starts at -1368 and continues past the end of the *facB* sequence (-1380) for a further 6 bp of polylinker sequence (-1386). This later window extends over the polylinker

region which consists of a fused *Bam* HI/*Sma* I site. The windows are 18 and 19 bp in length respectively, and are present when either concentration of DNase I was used. The third less well protected region of 8 bp spans the region from -104 to -111 and this corresponds to the sequence protected weakly on the non-coding strand. These results are summarised in Figure 3.16a.

As was the case with the ALU148 fragment from *amdS*, polylinker sequences were protected at one end of the FACB1 fragment by GST-CREA. Unlike the situation with the ALU148 fragment the presence of a restriction endonuclease site within the FACB1 fragment made it possible to remove the polylinker sequences and only 7 bp of the *facB* sequence for use in footprinting. The *Hind* III-*Eco* RI FACB1 fragment was labelled, digested with *Hinf* I restriction endonuclease and used in a footprinting experiment. However, this fragment could not be footprinted as it showed binding to control pGEX-2T extracts. A gel mobility shift assay (Fig. 3.16b) showed that removing 13 bp from the end of this fragment changed the binding dramatically, and the retardation seen in the GST-CREA containing track was identical to that in the control GST containing track. This was reproducible with different preparations of pGEX-2T extracts. It is not clear why binding to pGEX-2T control extracts was not seen in the larger FACB1 fragment.

3.6.1 Summary

Two regions spanned by the 239 bp *Nhe* I-*Sac* I and 273 bp *Sca* I- *Taq* I fragments, showed binding to CREA. In addition weak binding to the 219 bp *Taq* I-*Nru* I fragment cannot be ruled out. As binding to the 219 bp *Taq* I-*Nru* I and 239 bp *Nhe* I-*Sac* I fragments could not be localised to smaller sub-fragments, a footprinting analysis of these regions could not be carried out. An analysis of the sequence of these fragments may indicate why sub-fragments are no longer able to be bound by CREA. Footprinting of the weakly binding 130 bp *Nla* IV- *Taq* I fragment was not attempted. The FACB1 sub-fragment of the 273 bp *Sca* I- *Taq* I fragment, when used in footprinting, was able to bind CREA but a shorter *Hinf* I- *Eco* RI sub-fragment

Figure 3.15a DNase I sensitivity assay (footprinting) on the non-coding strand of FACB1

The FACB1 fragment from *facB*, labelled at the *Bam* HI end, was incubated with either 50 ng or 100 ng of DNase I in the presence (**R**) or absence (**F**) of the GST-CREA fusion protein and electrophoresed on a mobility shift gel. Bands corresponding to the free fragment (**F**) and the retarded complex (**R**) were excised and eluted. After purification samples were electrophoresed on a sequencing gel alongside a Maxam and Gilbert "G" sequencing reaction (**S**).

The sequence of the fragment pictured is shown on the left. Alongside the sequence the two protected regions of the **R** complex are shown as open boxes. Hyperdigested bases are marked with an * and summary of these findings is shown in Figure 3.16a.

The DNA fragments used for footprinting were a mixture of two sizes due to the presence of two *Nla* III sites close together, this did not affect footprinting on this strand.

Figure 3.15b DNase I sensitivity assay (footprinting) on the coding strand of FACB1

The FACB1 fragment from *facB*, labelled at the *Eco* RI end, was incubated with either 50 ng or 100 ng of DNase I in the presence (**R**) or absence (**F**) of the GST-CREA fusion protein and electrophoresed on a mobility shift gel. Bands corresponding to the free fragment (**F**) and the retarded complex (**R**) were excised and eluted. After purification samples were electrophoresed on a sequencing gel alongside a Maxam and Gilbert "G" sequencing reaction (**S**).

Upon drying the sequencing gel cracked and therefore not all bases were visible. Those that are, are shown in the sequence on the left. Alongside the sequence the three protected regions of the **R** complex are shown as open boxes. Hyperdigested bases are marked with an * and a summary of these findings is shown in Figure 3.16a.

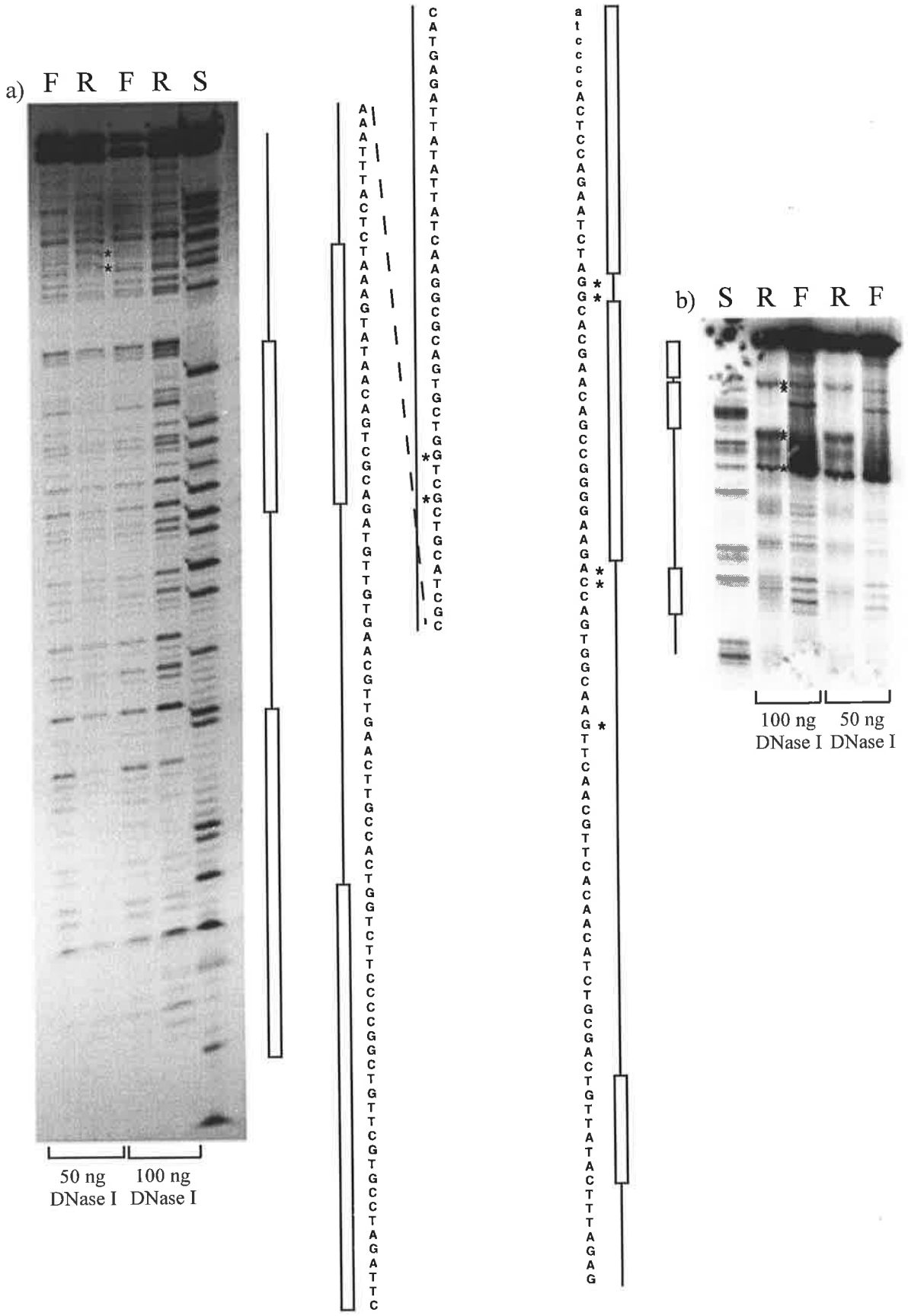

a)
F R F R S
50 ng DNase I
100 ng DNase I
b)
S R F R F
100 ng DNase I
50 ng DNase I

Figure 3.16a Summary of the protected regions of the FACB1 fragment

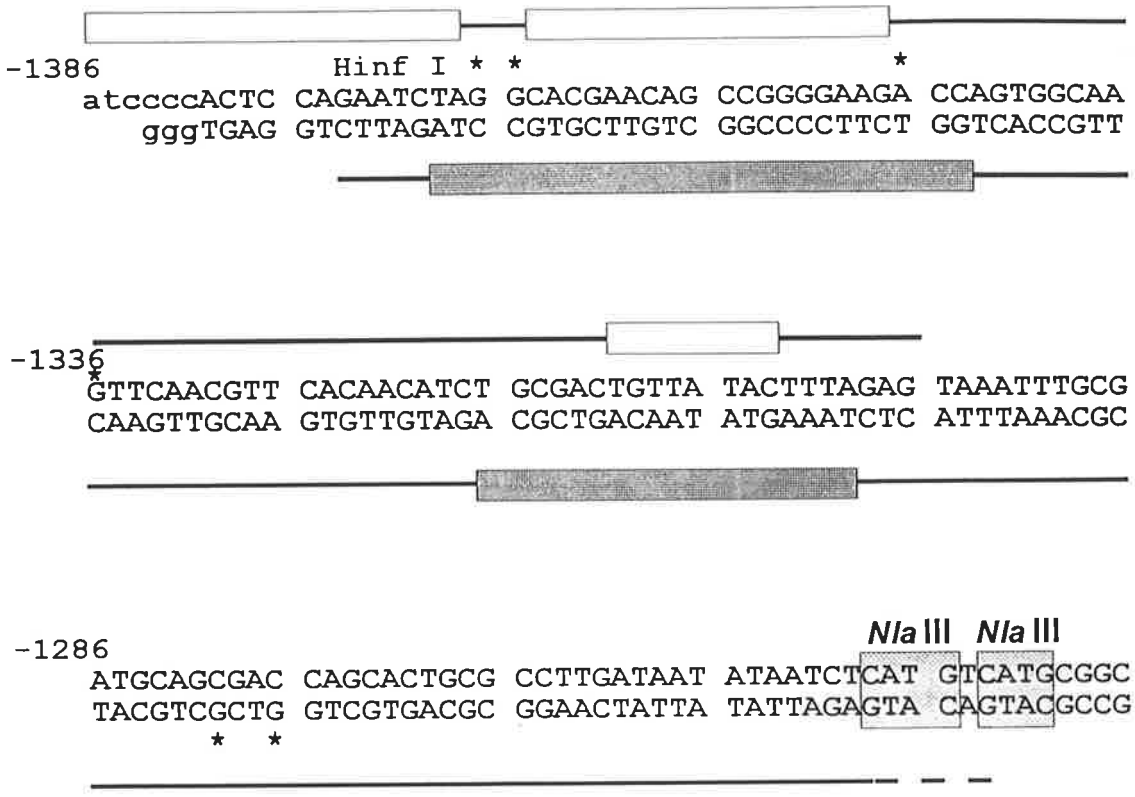
The sequence of the FACB1 fragment from *facB* including some restriction endonuclease sites and the nucleotide positions relative to the start point of translation are shown. The protected regions (boxed) on the coding strand are shown above the sequence and the protected regions for the non-coding strand are shown below the sequence. Hyperdigested bases are marked with an * .

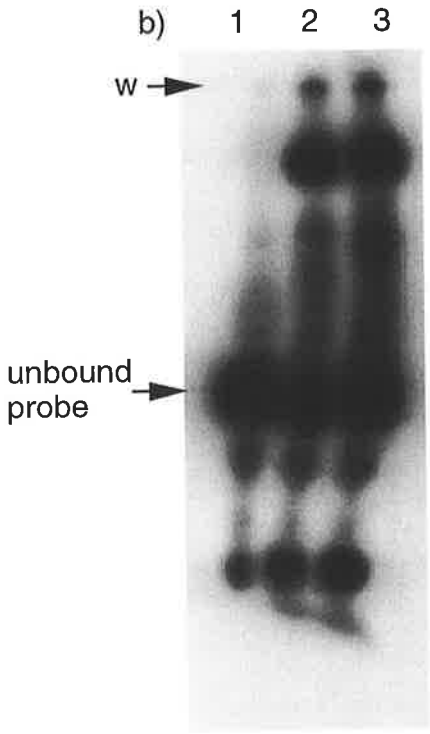
Figure 3.16b Mobility shift analysis of FACB1 digested with *Hinf* I

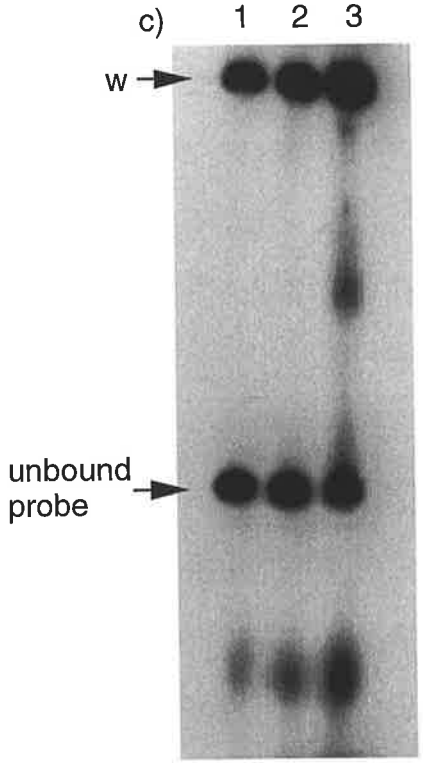
A mobility shift assay of the FACB1 fragment from *facB* digested with *Hinf* I restriction endonuclease and then incubated with PBS (**Track 1**) or 300 µg total protein extract containing GST (**Track 2**) or 160 µg total protein extract containing GST-CREA (**Tracks 3**).

Figure 3.16c Mobility shift analysis on the ONC12/13 PCR fragment from *facB*

A mobility shift assay of 10 ng of double stranded ONC12/13 PCR product, incubated with PBS (**Track 1**) or 300 µg total protein extract containing GST (**Track 2**) or 160 µg total protein extract containing GST-CREA (**Track 3**). The ONC12 primer contains a restriction site for *Bam* HI and ONC13 a site for *Eco* RI.


-1386
Hinf I * *
*
atccccACTC CAGAATCTAG GCACGAACAG CCGGGGAAGA CCAGTGGCAA
gggTGAG GTCTTAGATC CGTGCTTGTC GGCCCCTTCT GGTCACCGTT
-1336
*
GTTCAACGTT CACAACATCT GCGACTGTTA TACTTTAGAG TAAATTTGCG
CAAGTTGCAA GTGTTGTAGA CGCTGACAAT ATGAAATCTC ATTTAAACGC
-1286
NlaIII NlaIII
ATGCAGCGAC CAGCACTGCG CCTTGATAAT ATAATCTCAT GTCATGCGGC
TACGTCGCTG GTCGTGACGC GGAACTATTA TATTAGAGTA CAGTACGCCG
* *


b)
1
2
3
w
unbound
probe


c)
1
2
3
w
unbound
probe

was not. These results will be discussed further in section 3.9.3 where a mobility shift analysis of the ONC12/ONC13 PCR product (Fig. 3.16c) will also be presented.

3.7 Mobility shift analysis of the 5' *creA* region

In order to investigate whether a possible autoregulatory role for the CREA gene product could be a result of direct regulation by CREA, the 5' region of the *creA* gene was tested for binding to the GST-CREA fusion protein. A partial restriction endonuclease map of the 5' region of *creA* and the DNA fragments which were used in mobility shift analyses are shown in Figure 3.17a. Two large fragments, *Bam* HI-*Ava* I and *Ava* I-*Pst* I, were isolated and digested with *Hinf* I to produce five sub-fragments. These were used in mobility shift assays with the GST-CREA fusion protein. The results (Fig. 3.17b) show binding by the GST-CREA fusion protein to the 243 bp *Bam* HI-*Hinf* I and 127 bp *Ava* I-*Hinf* I fragments but not to the other three fragments.

3.8 Determining the recognition sequence of GST-CREA

Nardelli *et al.* (1991) compared the zinc fingers of the Krox-20 (Chavrier *et al.*, 1988) and Sp1 (Kadonaga *et al.*, 1987) proteins to their target sites and produced a model for the interaction of zinc fingers with DNA. They hypothesised that each finger recognises a triplet of bases and the sequence recognised was dependent on the amino acids present at specific key locations of the zinc finger, namely positions -1, +3 and +6. Nardelli *et al.*, (1991) showed that an R(+6)E(+3)R(-1) zinc finger will recognise a GCG triplet and THR, KHR and AHK zinc fingers recognise a GGG triplet (Fig. 3.18a). The specificity of a finger could be changed by mutation. For example, mutation of the Krox-20 zinc finger II from THR to RER changed the binding specificity from a GGG triplet to a GCG triplet. Jacobs (1992) searched the data base of classical zinc fingers looking for positions where amino acids were similar or highly divergent and also deduced that the crucial amino acids involved in DNA binding were amino acids -1, +3 and +6 (s3, s6 and m3 respectively using Jacobs' notation).

These results were confirmed by crystallographic analysis of Zif268 (Paveletich and Pabo, 1991). Three amino acids on the exposed face of the α-helix made base specific contacts, residue -1 immediately preceding the α-helix and residues +3 and +6 within the α-helix. For a number of zinc fingers the recognition sites and the amino acids at the three critical positions are also shown in Figure 3.18a. For Zif268, R(-1) and R(+6) make two specific contacts with a guanine and H(+3) makes one contact with a guanine. These are the only base specific contacts made by Zif268 (Paveletich and Pabo, 1991). Surprisingly not every base in the recognition site was specified through contacts with amino acid side chains. It is not known for example how the cytidine in the triplet GCG is specified.

3.8.1 The consensus binding site for MIG1

The MIG1 repressor from *S. cerevisiae* and CREA have very similar zinc fingers (85% identity and 96% similarity at the amino acid level) and at the three critical residues MIG1 and CREA are identical, having RER (II) and RHR(I) fingers. It is therefore likely that CREA and MIG1 would recognise very similar core sequences. From the studies outlined above, the predicted binding site for MIG1 would be 5' GCG GGG 3' (where the 3' G is referred to as G1). Using *in vitro* transcribed and translated protein, MIG1 was shown to bind to two sites 5' of the *SUC2* gene (Nehlin and Ronne, 1990). These sites 5' of *SUC2* and *GAL4* lie within regions defined as essential for carbon catabolite repression of these genes, URS_G (upstream repression sequences, Flick and Johnson, 1990; Griggs and Johnson, 1991). The **A** site in *SUC2* (Nehlin and Ronne, 1990; Fig. 3.18b) contains a GCG GGG sequence identical to that predicted above. The **B** site is located 40 bp closer to the start point of transcription and contains a sequence which differs at one position from the predicted sequence (ie CCG GGG), and thus is missing an arginine-guanine contact (if the results of crystallographic studies on Zif268 are applicable to MIG1). Nehlin *et al.*, (1991) looked at the 5' region of two other genes, *GAL1* and *GAL4*, and found that while the GGG was present in all binding sites, the broader consensus for the second finger was SYG. In addition the MIG1 protein has an extended consensus of 5' WWWWT 3' which is both protected in DNase I footprints and present in all natural MIG1 binding sites. A

Figure 3.17a Partial restriction endonuclease map of the 5' region of *creA*

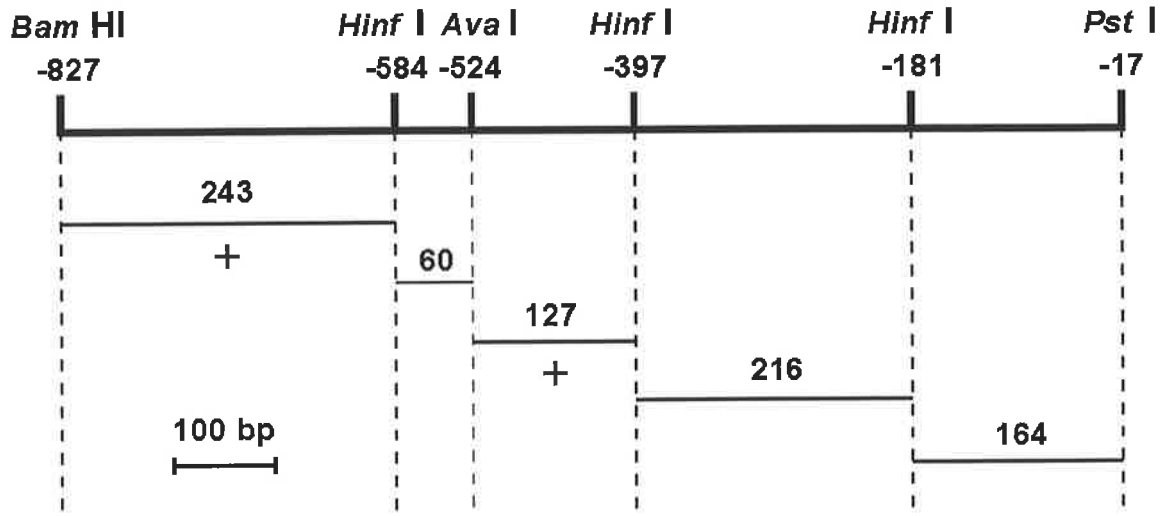
Scale map of the 5' region of the *creA* gene (Dowzer and Kelly, 1991; Shroff *et al.*, submitted). Fragments isolated for mobility shift assays are shown as horizontal bars with sizes marked above the bars. Fragments are referred to in the text by size and by the names of the restriction endonucleases used to isolate them. Fragments which bind the GST-CREA fusion protein are marked with a +. The start point of translation is +1.

Figure 3.17b Mobility shift analysis of the 5' region of *creA*

A mobility shift assay of the 243 bp and 60 bp fragments from the *Bam* HI- *Ava* I region and the 127 bp, 216 bp and 164 bp fragments from the *Ava* I- *Pst* I region 5' of *creA*.

In all cases the first Track contains labelled fragment incubated with PBS, the second with 300 μg total protein extract containing GST and the third Track with 160 μg total protein extract containing GST-CREA.

a)


Bam HI
-827
Hinf I
-584
Ava I
-524
Hinf I
-397
Hinf I
-181
Pst I
-17
243
+
60
127
+
216
164
100 bp

b)

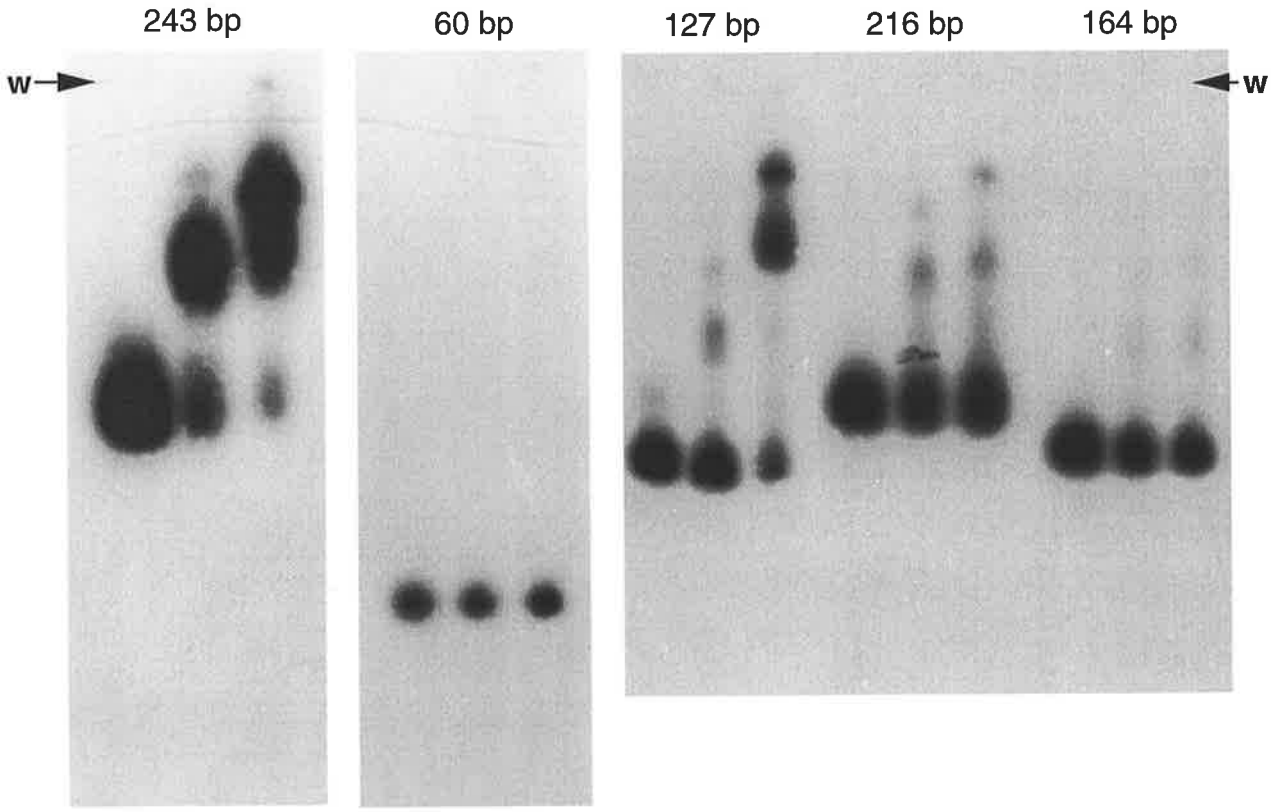

243 bp
60 bp
127 bp
216 bp
164 bp
w
w

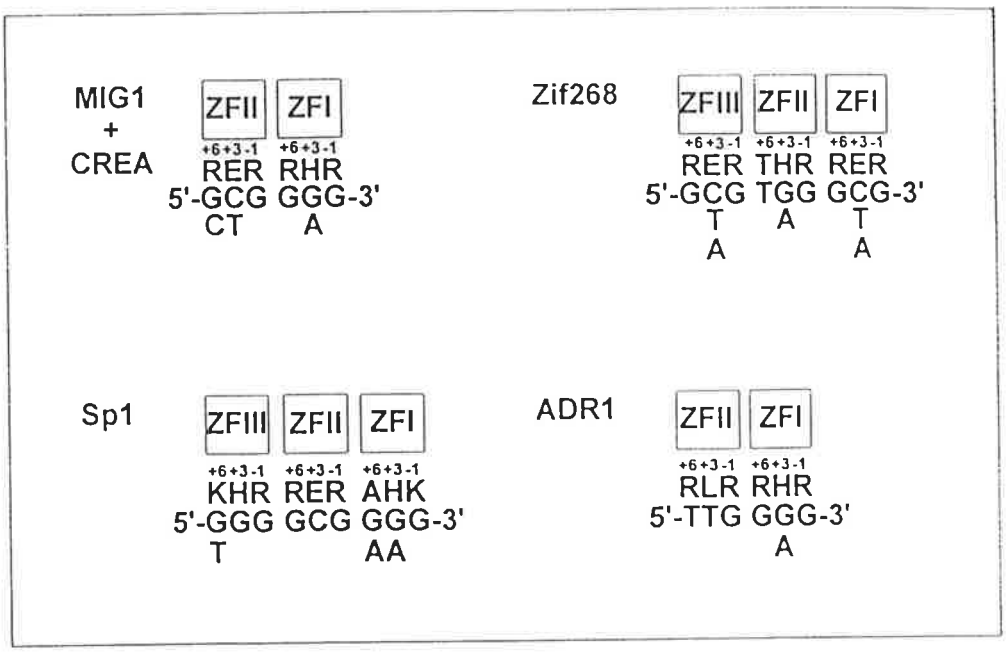
Figure 3.18a Schematic diagram of the interaction of zinc fingers with their recognition sequence

For some Cys_2His_2 zinc finger proteins the amino acid residues present at positions +6 (21), +3 (18) and -1 (15) of the zinc finger motif and the bases in the recognition sequence that they are able to specify is shown. This diagram is a modification of Figure 7 from Lundin *et al.*, (1994). The CREA and MIG1 proteins have identical amino acid residues at these positions.

Figure 3.18b Consensus binding site for the MIG1 protein from *S. cerevisiae*

The MIG1, DNase I protected regions from the 5' regions of the *SUC2*, *GAL1* and *GAL4* genes (Nehlin *et al.*, 1991) is shown. The consensus binding site is shown in bold.

a)



b)

		protected bases

SUC2	A	AGTAATAAAAATGCGGGGAAT
SUC2	B	TTAGGAAATTATCCGGGGGCG
GAL4		GAAGCTGAAAATCTGGGGAAG
GAL1		TTAGCCTTATTTCTGGGGTAA
CONSENSUS		WWWWTSYGGGG

saturated mutagenesis of the naturally occurring *SUC2* **A** site was undertaken to further elucidate its binding specificity (Lundin *et al.*, 1994). Three deviations from GCG GGG retained binding activity; GCG GAG, CCG GGG and GTG GGG. In addition oligonucleotides containing two of these changes retained binding activity and to a lesser extent so did an oligonucleotide containing all three changes, ie CTG GAG. Although a T7 is present in all naturally occurring MIG1 binding sites, changing this to a C or A only mildly reduced binding and changing to a G had no obvious effect on binding affinity. Within the extended consensus of 5' WWWW 3' (the AT box) specific bases were not important for binding but rather the overall AT richness was important. If more than one C residue was present in this AT box, binding by MIG1 was abolished. A DNA helix containing this AT rich region is flexible, and Lundin *et al.*, (1994) proposed that this bending is required for MIG1 binding. The model presented has as a first step binding of MIG1 via its zinc fingers to the GC box (low affinity) and subsequent bending of the target site to allow binding to the AT box by a different portion of the MIG1 protein (high affinity).

3.8.2 Sequence analysis of the ALU148 fragment and mobility shift assays using oligonucleotides from the protected region of *amdS*

Footprinting on the ALU148 fragment showed that the coding strand was protected in two windows which were also protected on the non-coding strand. These two windows are likely to contain the recognition site(s) for CREA since they were protected on both strands. These sequences are located just upstream of the start point of transcription which occurs at -41 bp (Fig. 3.19). The region -81 to -122 bp is GC rich (67%: GC(28): AT(14)) and contains three GGGG sequences. The small regions protected by CREA span approximately 17-20 bp, and therefore it seems likely that the large window of approximately 62 bp, which contains these three GGGG sequences, could be expected to contain multiple sites. **A1** and **A4** are the same as the MIG1 consensus, SYG GGG and **A2** differs at one position GAG GGG. **A3** does not have 4G's but differs at only one position to SYG GRG ie GAG GAG and was protected only on the non-coding strand. The **A1** sequence has the G rich strand as the non-coding strand

which may be significant given that the 5' region of SUC2 contains one string of 4C's (A site) and one of 4G's (B site) on the coding strand. The region protected by CREA is quite large and therefore an oligonucleotide approach was used to identify a smaller sequence which was sufficient for binding.

Two pairs of complimentary oligonucleotides, 31/ 32 and RL2/RL3 (Table 2.3), spanning the region 5' of *amdS* (Fig. 3.19) were used in a gel mobility shift assay and the result is shown in Figure 3.20a. There was only weak binding by GST-CREA to either oligonucleotide pair and thus these sequences were not sufficient for recognition and strong binding by CREA. An oligonucleotide ONC1 (Fig. 3.19) was synthesised (Table 2.3) and made double stranded in the presence of labelled dATP and unlabelled dCTP, dTTP and dGTP using ONC7 as a primer. The resulting double stranded oligonucleotide, when used in a gel mobility shift assay, was found to bind strongly to the CREA fusion protein (Fig. 3.20b). Therefore ONC1/ONC7 contains sufficient sequences for recognition and strong binding by CREA.

To test the importance of specific bases present in the ONC1/ONC7 oligonucleotide a series of oligonucleotides (ONC4, 5, 6, 10) with the sequence changes shown in Figure 3.19 were made. These oligonucleotides were made double stranded with ONC7, labelled and used in gel mobility shift assays. The results show that ONC10/ONC7 bound CREA, ONC4/ONC7 and ONC6/ ONC7 did not bind and ONC5/ONC7 bound weakly (Fig. 3.20b). These results will be discussed below in section 3.9.

3.9 Discussion

The Cys_2His_2 DNA binding motif present in the theoretically translated amino acid sequence of CREA was shown to be active *in vitro* by mobility shift and DNase I sensitivity assays. The *E. coli* expressed fusion protein GST-CREA was able to bind DNA in a sequence specific manner and this binding was independent of other *A. nidulans* proteins. As expected for zinc finger binding domains, DNA binding was dependent on the presence of Zn^{2+} ions.

Figure 3.19 Sequence of the protected region from *amdS* and the synthetic oligonucleotides used in this study

The position of the protected region (coding strand) and the synthetic oligonucleotides 31/32, RL2/RL3 and ONC1/ONC7, ONC4/ONC7, ONC5/ONC7, ONC6/ONC7 and ONC10/ONC7 used in this study, is shown. The arrows represent changes to the wild type *amdS* sequence (ONC1/ONC7). The shaded boxes highlight the GC rich sites present in this region. The start point of translation is +1.

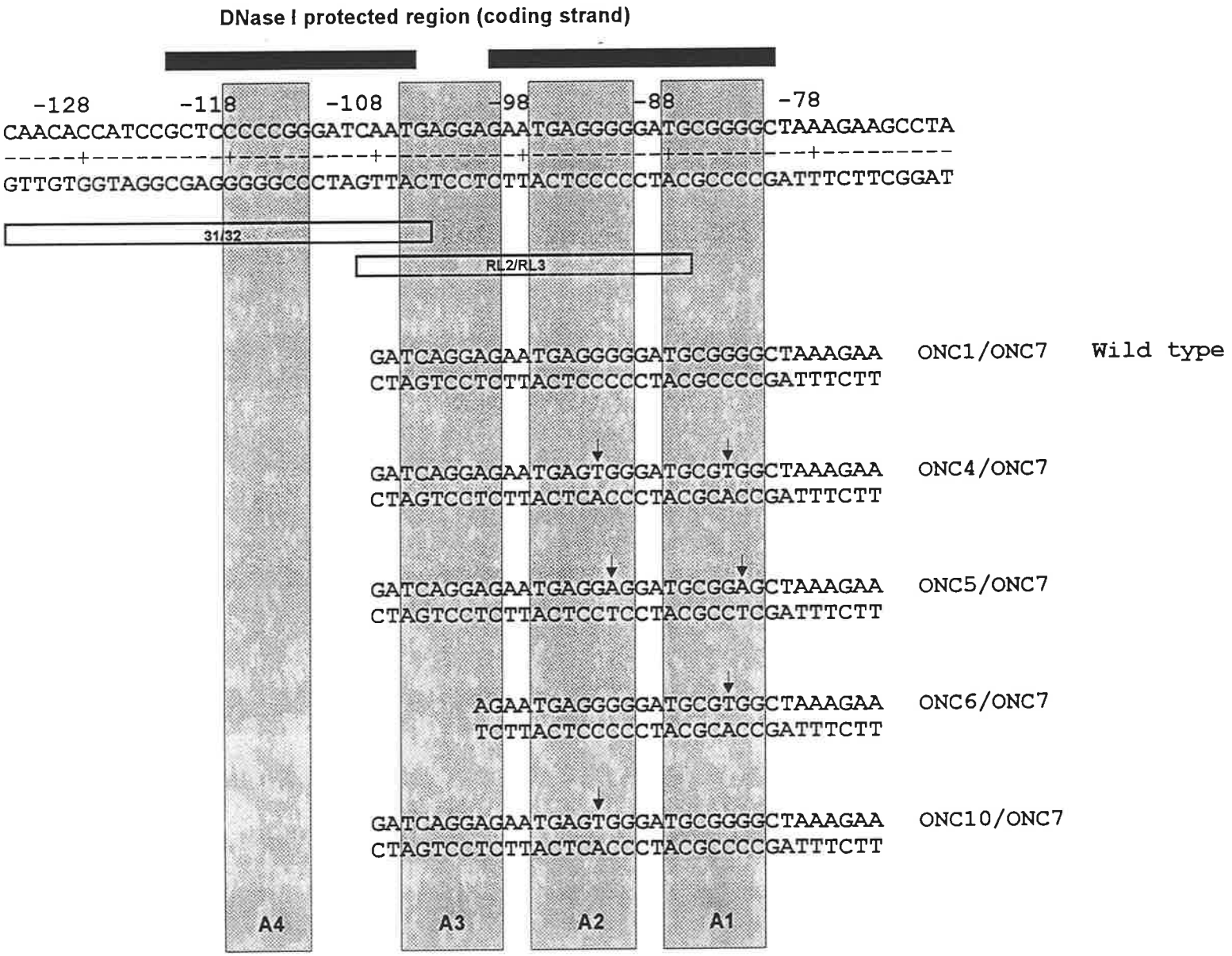

DNase I protected region (coding strand)
-128
-118
-108
-98
-88
-78
CAACACCATCCGCTCCCCCGGGATCAATGAGGAGAATGAGGGGGATGCGGGGCTAAAGAAGCCTA
GTTGTGGTAGGCGAGGGGGCCCTAGTTACTCCTCTTACTCCCCCTACGCCCCGATTTCTTCGGAT
31/32
RL2/RL3
GATCAGGAGAATGAGGGGGATGCGGGGCTAAAGAA
CTAGTCCTCTTACTCCCCCTACGCCCCGATTTCTT
ONC1/ONC7
Wild type
GATCAGGAGAATGAGTGGGATGCGTGGCTAAAGAA
CTAGTCCTCTTACTCACCCTACGCACCGATTTCTT
ONC4/ONC7
GATCAGGAGAATGAGGAGGATGCGGAGCTAAAGAA
CTAGTCCTCTTACTCCTCCTACGCCTCGATTTCTT
ONC5/ONC7
AGAATGAGGGGGATGCGTGGCTAAAGAA
TCTTACTCCCCCTACGCACCGATTTCTT
ONC6/ONC7
GATCAGGAGAATGAGTGGGATGCGGGGCTAAAGAA
CTAGTCCTCTTACTCACCCTACGCCCCGATTTCTT
ONC10/ONC7
A4
A3
A2
A1

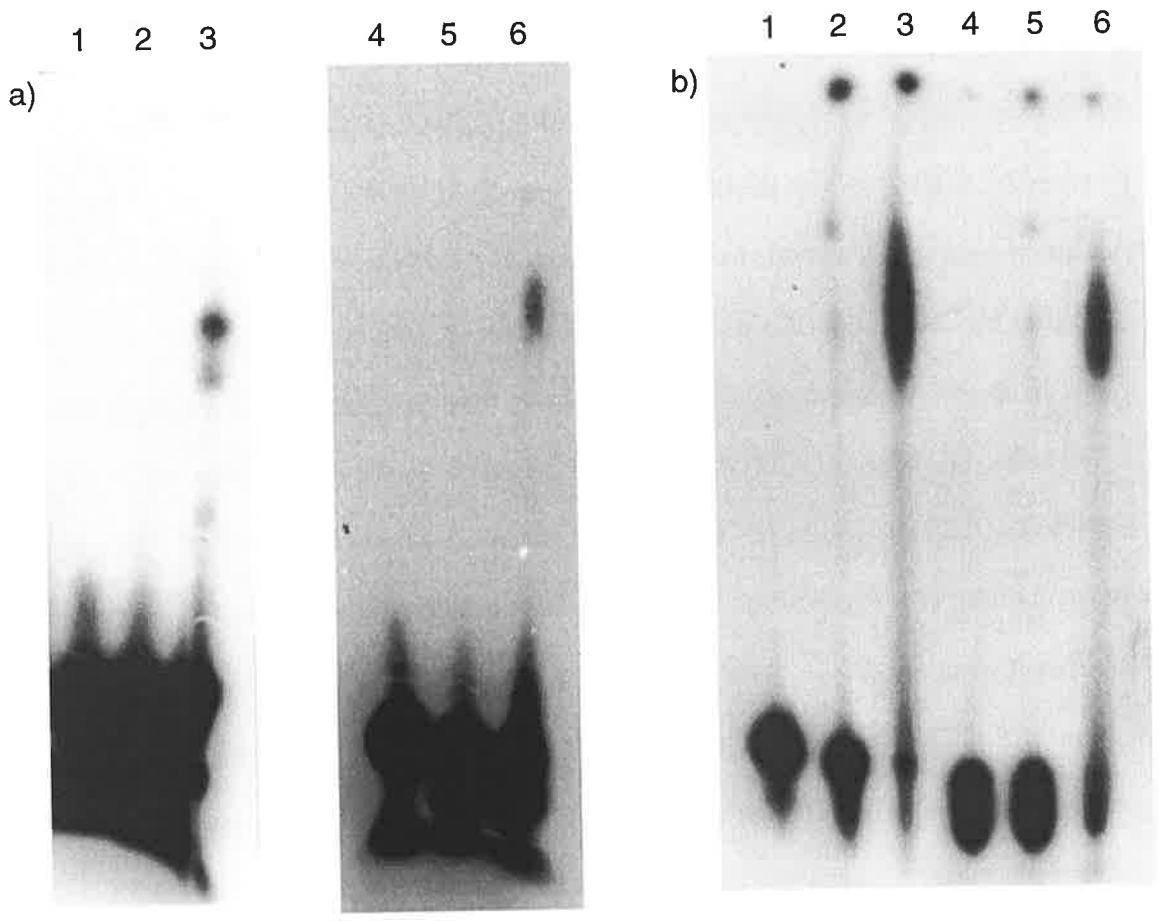
Figure 3.20a Mobility shift analysis of the synthetic oligonucleotides 31/32 and RL2/RL3 from the 5' region of the *amdS* gene

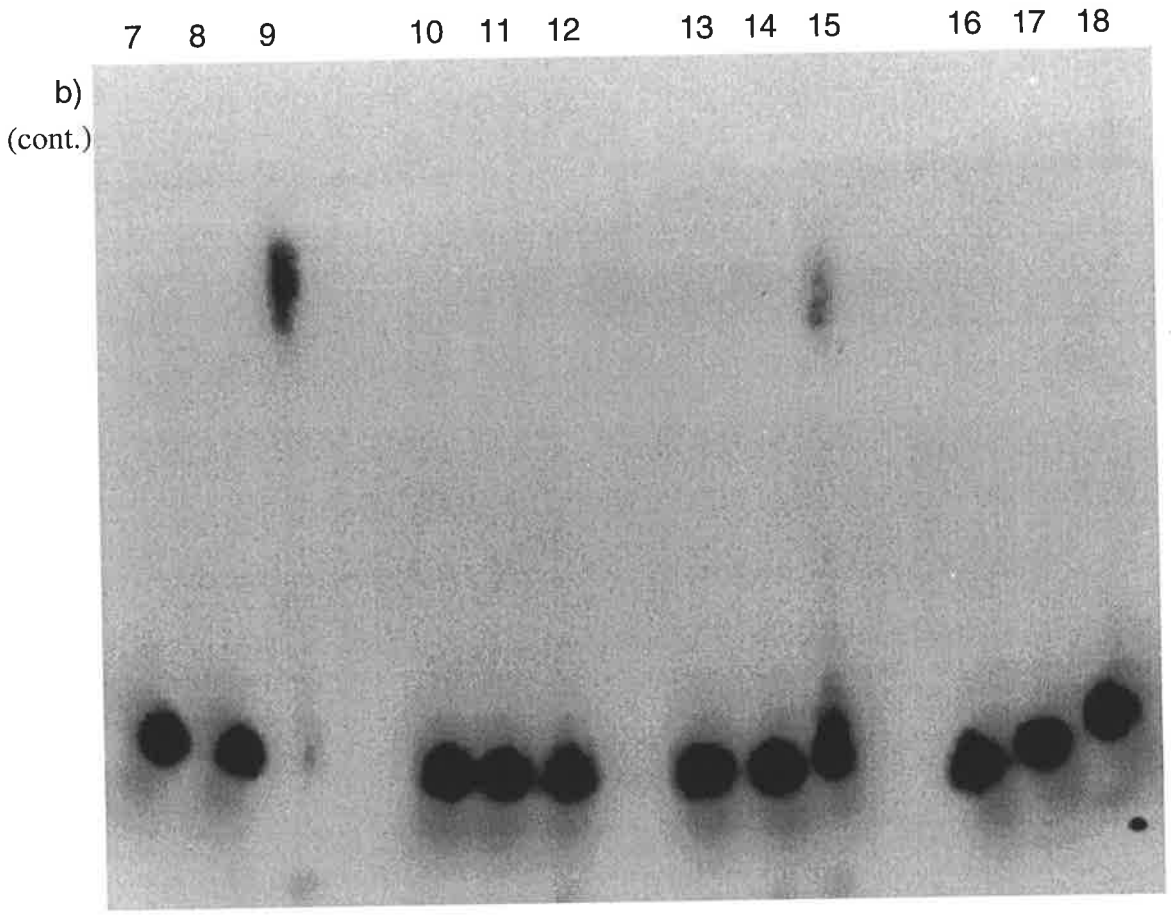
Oligonucleotides 31, 32 and RL2 and RL3 were kindly supplied by M.J. Hynes. The oligonucleotide pairs were each annealed in 1X sequenase buffer and end labelled with the Klenow fragment of DNA polymerase I. The double stranded oligonucleotides were used in a mobility shift assay. Pictured are the oligonucleotide pairs 31/ 32 (**Tracks 1-3**), and RL2/RL3 (**Tracks 4-6**).

Figure 3.20b Mobility shift analysis of ONC1/ONC7 and other synthetic oligonucleotides spanning the 5' region of the *amdS* gene

A mobility shift assay of double stranded synthetic oligonucleotides (Table 2.3) containing either the wild type *amdS* sequence (ONC1/ONC7, **Tracks 1-3** and **7-9**), two G3->T3 transitions within **A1** and **A2** (ONC4/ONC7, **Tracks 10-12**), two G2->A2 transitions within **A1** and **A2** (ONC5/ONC7, **Tracks 13-15**), a G3->T3 transition within **A1** (ONC6/ONC7, **Tracks 16-18**) or a G3->T3 transition within **A2** (ONC10/ONC7, **Tracks 4-6**) .

In all cases the first Track contains labelled fragment incubated with PBS, the second with 300 μg total protein extract containing GST and the third Track with 160 μg total protein extract containing GST-CREA.


a)
1 2 3 4 5 6
b)
1 2 3 4 5 6


b)
(cont.)
7 8 9 10 11 12 13 14 15 16 17 18

The oligonucleotide pair ONC1/ONC7 binds CREA strongly. This sequence contains three GC boxes. **A1** is identical to the functional *SUC2* **A** site bound by MIG1 and **A2** and **A3** only deviate from the MIG1 consensus at the fifth position. The single change within **A1** from G3 to T3 (ONC6/ONC7) and the double change within ONC4/ONC7 abolishes binding by the CREA fusion protein. This confirms the importance of the string of four G residues (4G string) in recognition by CREA. ONC6/ONC7 is however, a shorter oligonucleotide than ONC1/ONC7 as it also lacks the **A3** sequence. The lack of strong binding by RL2/RL3 suggests that two imperfect sites (**A2** (T GAG GGG) and **A3** (T CAG GAG)) are not sufficient for strong binding. A single SYG GGG site is also not sufficient since 31/32 did not bind strongly. This may be due to the absence of a T in position 7 or perhaps a wider context outside the GC box exists for CREA like it does for MIG1. Although binding to the ONC10/ONC7 oligonucleotide was strong the GST-CREA fusion protein appeared to have a higher affinity for the ONC1/ONC7 oligonucleotide. The only difference between the two oligonucleotides is that ONC10/ONC7 contains the **A1** and **A3** sites but lacks the **A2** site. The **A2** site differs from the SYG GGG MIG1 consensus at one position and thus it may be that a non consensus site is able to increase binding affinity *in vitro*.

The consensus for MIG1 contains a T7 and there is a T in the 7th position within **A1**, **A2** and **A3** sites but no obvious AT rich sequences directly 5' of this. It is not clear from studies on other zinc fingers how a T in the seventh position can be specified by the zinc finger motif. Perhaps the T is involved in interactions with amino acid residue(s) outside the second finger or perhaps elsewhere in the protein. Since there is very little sequence similarity between MIG1 and CREA outside of the zinc finger region there is no reason to expect that a T7 or an AT rich region 5' of the recognition site is required by CREA. The conservative change from G2 to A2 within **A1** and **A2** results in some binding consistent with the results of Lundin *et al.*, (1994) that MIG1 was able to recognise a GCG GAG core sequence with a lower affinity than the GCG GGG consensus. In the following chapter the question of preferences within the core sequence, at the seventh position and 5' and 3' extended context will be addressed using a degenerate oligonucleotide approach.

The protection of the RL2/RL3 and 31/32 sequences within the larger ALU148 fragment and the fact that a number of very similar sequences are present in close proximity to the start point of transcription suggests that more than one molecule of CREA may be able to bind and that this binding may be co-operative. Interaction between more than one GST-CREA molecule may be stabilised *in vitro* by the GST moieties since GST exists as a dimer *in vivo* (Mannervik, 1985). It can be seen from Figure 3.3c that the fusion protein forms at least three complexes with the ALU148 fragment. Separate footprinting of the HMC and the LMC, on the non-coding strand, shows that the protected regions were very similar. This contradicts the hypothesis that in the LMC additional binding sites are occupied. Perhaps the only difference between the two complexes is their conformation. Alternatively, it is possible that dimerisation is occurring but that this does not involve binding to additional sites. The ALU148 fragment used in footprinting contains an artificial string of 4Gs created from the fusion of the *Alu* I- *Alu* I fragment and the *Sma* I site of pUC19. This region was protected on one strand only, and what effect this sequence has on binding is not clear. SA104 is able to bind CREA and lacks the artificial binding site present in ALU148. Only one complex was seen with the FACB1 fragment despite the presence of three windows in the protected region. An artificial GTG GG(G) sequence was present at the end of the FACB1 fragment and was protected in footprints. What effect this sequence had on footprinting is also not known.

A spacing of 2 bp between sites **A1**, **A2** and **A3** is equal to a binding site being repeated (twice) every nine base pairs. This is almost a complete turn of a β-DNA helix which implies that if two molecules of CREA are binding adjacent sites (and there is no real evidence for this at this stage) then they may be on the same side of the helix.

3.9.1 *In vivo* analysis of strains carrying mutations in the CREA boxes

The results presented in this chapter describe the *in vitro* binding of CREA to the 5' upstream region of *amdS, facB* and *creA*. The *in vivo* work on *amdS* has been undertaken in the laboratory of M.J. Hynes as part of on going research into the co-ordinated regulation of the *amdS* gene. A plasmid was constructed where the two putative recognition sequences **A1** and **A2**, shown

here to be important for binding *in vitro,* were deleted. DNA containing this deletion was placed upstream of the *lacZ* reporter gene and in a two step gene replacement strategy (Davis *et al.*, 1988) the wild type *amdS* gene was replaced with the 5' *amdS-lacZ* construct. It was found that a strain containing such a deletion was derepressed for β-galactosidase activity but to only half the level observed in a *creA204* mutant (M.J. Hynes pers. comm.).

This group also made 5' *amdS-lacZ* fusion constructs with mutations in one or other of the **A1** and **A2** sites and assayed these for β-galactosidase activity. It was found that these mutations all gave approximately the same level of derepression in the presence of glucose and that this was less than the level of derepression seen for the deletion (M.J. Hynes pers. comm.).

This information supports the results presented here that CREA recognises more than one site. However, the *in vivo* result that mutation of either site results in derepression suggests that the **A1** and **A2** binding sites play an equal role in repression despite the fact that **A2** deviates from the MIG1 consensus and might therefore be expected to be less important. Perhaps repression is a measure of the number of binding sites for CREA and that strongly repressed genes have more CREA binding sites than those that are not as strongly repressed.

It is possible that all the sites 5' of *amdS* contribute to control via CREA including site **A4**, containing the *Sma* I site. When tested for binding, a fragment containing the *amdI9* mutation and a *Bam* HI linker inserted into the *Sma* I site was able to bind CREA and do so with an affinity similar to the wild type fragment of *amdS* (Fig. 3.11). Minor changes in binding affinity would not be detected using the large excess of fusion protein as was used in this study. However, binding to this fragment would not be expected to be disrupted since many potential binding sites, including **A1** and **A2** were still intact. The *amdI9* mutation does not fall within a sequence with similarity to the MIG1 binding site and thus the only site expected to be disrupted within this fragment was the **A4** site. A fragment containing the *amdI18* mutation was also found to be unaffected for binding by CREA. The *amd I18* mutation is a single base change

within the **A4** site. Once again other potential binding sites are intact within this fragment and thus binding would not be expected to be disrupted.

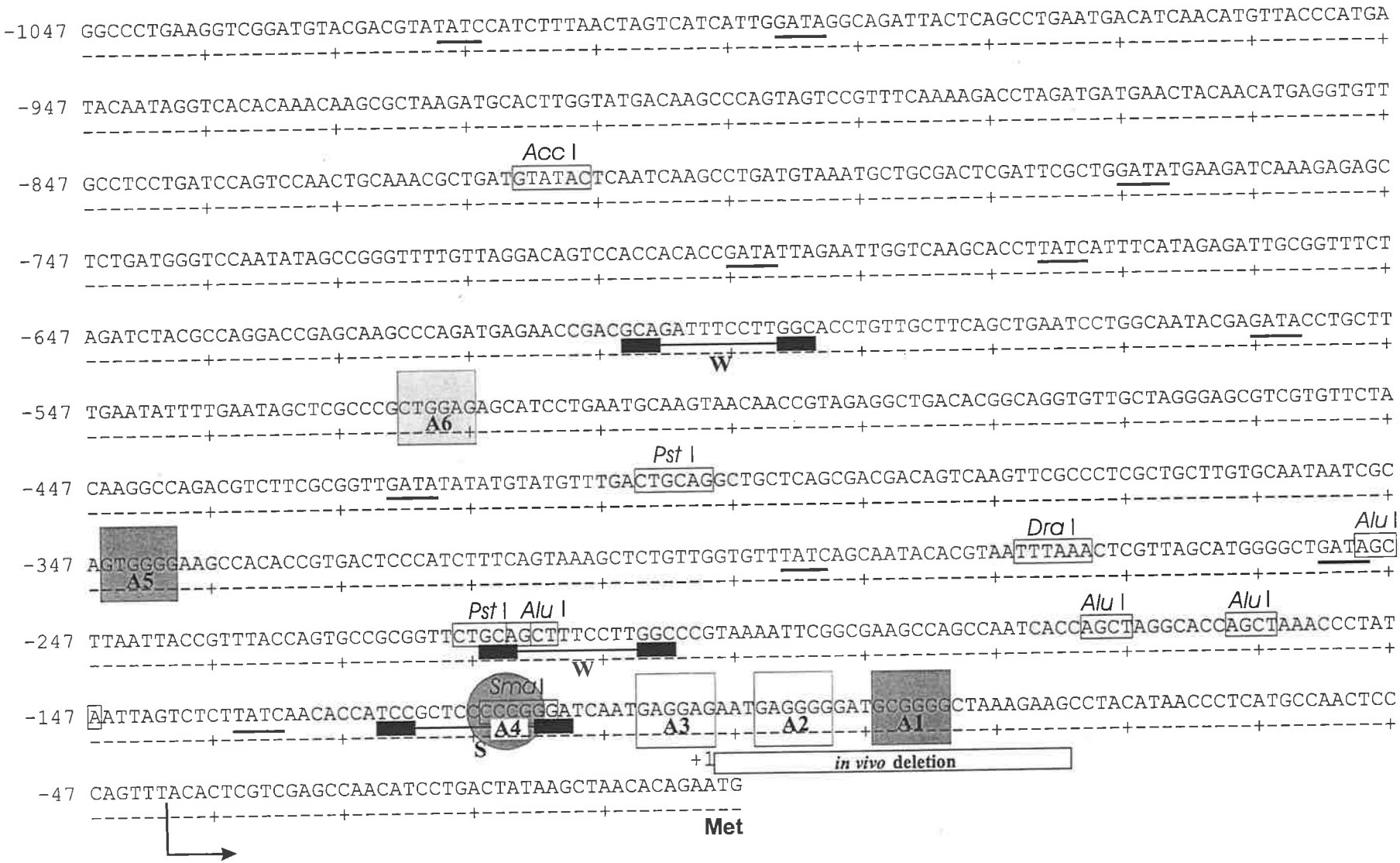
3.9.2 Recognition sites present in the 5' sequence of *amdS*

The sequence of the 5' region of *amdS* is presented in Figure 3.21 and is not contiguous with the sequence of the *Sal* I-*Sph* I region shown in Figure 3.7. Three SYG GGG sequences are present 5' of *amdS*, **A1**, **A4** and **A5**. Only one of these sequences (**A5**) lies outside of ALU148. **A5** lies within the PST186 fragment which showed retardation with CREA. When the DRA127 sub-fragment, which contains the **A5** sequence, was tested it showed only slight retardation that was not as great as that seen for the larger PST186 fragment. Within the larger PST186 fragment there are no other sequences that are similar to the SYG GGG consensus for MIG1 binding. Why a single site within a large fragment will bind but the same site within a smaller fragment will not is not known. Perhaps CREA is able to recognise another unrelated recognition sequence that was present in the larger but not the smaller fragment. As found in section 3.8.2, oligonucleotides from the 5' region of *amdS* containing just a single site did not bind CREA strongly. An oligonucleotide containing the **A1**, **A2** and **A3** sites was sufficient for strong binding, implying that the reason the DRA127 fragment did not bind CREA was perhaps because another site(s) was not present.

In addition to the sites already mentioned, a sequence fitting the broader consensus SYG GRG is present 5' of *amdS* (**A6**, Fig. 3.21). A large fragment containing the **A6** sequence did not bind GST-CREA. This suggests that perhaps the **A6** site represents a weaker site for CREA and that even within a large fragment this site is not sufficient to allow binding. The *in vivo* work (section 3.9.1) suggests that the ONC1/ONC7 region is not sufficient to entirely explain the carbon catabolite repression at the *amdS* locus. It is possible that a site upstream can interact with proteins bound near the start point of transcription by looping out of the intervening DNA and that these distant sites can play a role in carbon catabolite repression.

Figure 3.21 Sequence of the 5' region of *amdS*

The sequence of the the 5' region of the *amdS* gene and some restriction endonuclease sites are shown (Corrick *et al.*, 1987). All sequences fitting the consensus SYG GGG (dark shading), the broader consensus SYG GRG (light shading) or which deviate from the broad consensus but are protected (unshaded) are shown and are labelled **A1** to **A6**. The orientation of the sites is shown by squares (G rich strand is the coding strand) or circles (G rich strand is the non-coding strand). The location of the *in vivo* deletion (M.J. Hynes pers. comm.) is shown with a rectangle. Potential binding sites for the AREA protein (GATA/ TATC sequences; Merika and Orkin, 1993) on both strands of the DNA are underlined and the position of FACB binding sites (S=strong, W= weak) is also shown (Todd, 1995). These will be discussed in chapter 6. The start point of translation is +1.


-1047 GGCCCTGAAGGTCGGATGTACGACGTATATCCATCTTTAACTAGTCATCATTGGATAGGCAGATTACTCAGCCTGAATGACATCAACATGTTACCCATGA
-947 TACAATAGGTCACACAAACAAGCGCTAAGATGCACTTGGTATGACAAGCCCAGTAGTCCGTTTCAAAAGACCTAGATGATGAACTACAACATGAGGTGTT
Acc I
-847 GCCTCCTGATCCAGTCCAACTGCAAACGCTGATGTATACTCAATCAAGCCTGATGTAAATGCTGCGACTCGATTCGCTGGATATGAAGATCAAAGAGAGC
-747 TCTGATGGGTCCAATATAGCCGGGTTTTGTTAGGACAGTCCACCACACCGATATTAGAATTGGTCAAGCACCTTATCATTTCATAGAGATTGCGGTTTCT
-647 AGATCTACGCCAGGACCGAGCAAGCCCAGATGAGAACCGACGCAGATTTCCTTGGCACCTGTTGCTTCAGCTGAATCCTGGCAATACGAGATACCTGCTT
W
-547 TGAATATTTTGAATAGCTCGCCCGCTGGAGAGCATCCTGAATGCAAGTAACAACCGTAGAGGCTGACACGGCAGGTGTTGCTAGGGAGCGTCGTGTTCTA
A6
Pst I
-447 CAAGGCCAGACGTCTTCGCGGTTGATATATATGTATGTTTGACTGCAGGCTGCTCAGCGACGACAGTCAAGTTCGCCCTCGCTGCTTGTGCAATAATCGC
Dra I
Alu I
-347 AGTCGGGAAGCCACACCGTGACTCCCATCTTTCAGTAAAGCTCTGTTGGTGTTTATCAGCAATACACGTAATTTAAACTCGTTAGCATGGGGCTGATAGC
A5
Pst I
Alu I
Alu I
Alu I
-247 TTAATTACCGTTTACCAGTGCCGCGGTTCTGCAGCTTTCCTTGGCCCGTAAAATTCGGCGAAGCCAGCCAATCACCAGCTAGGCACCAGCTAAACCCTAT
W
Sma I
-147 AATTAGTCTCTTATCAACACCATCCGCTCCCCCGGGATCAATGAGGAGAATGAGGGGGATGCGCGGCTAAAGAAGCCTACATAACCCTCATGCCAACTCC
A4
A3
A2
A1
S
+1
in vivo deletion
-47 CAGTTTACACTCGTCGAGCCAACATCCTGACTATAAGCTAACACAGAATG
Met

Within the *Sal* I-*Sph* I sequence there are no SYG GGG sequences but there is a region (boxed in Fig. 3.7) which contains three sites that differ at only one position from SYG GRG. Two of these are separated by a distance of 2 bp (as was the case for **A1**, **A2** and **A3** from ALU148). It is possible that the CREA sites within the *Sal* I-*Sph* I region may be involved in the regulation of a gene transcribed in the opposite direction to *amdS*. Present in this sequence are GATA, CCAAT and TATA-like sequences (Fig. 3.7), however, a search of the EMBL data base (September, 1996) with the open reading frame present at the beginning of this sequence found no significant sequence similarity with other proteins.

3.9.3 Recognition sites present in the 5' sequence of *facB*

The sequence of the 5' *facB* region is shown in Figure 3.22. The two consensus SYG GGG sites **F5** and **F6** and the SYG GRG sites, **F1-F4** and **F7**, are shown. The fragment FACB1 contains a SYG GGG sequence (**F6**) and bound GST-CREA strongly. Footprinting analysis of FACB1 shows that in addition to the **F6** sequence another sequence **F7**, CTC CAG (CTG GAG on the non-coding strand) was also protected by GST-CREA. When the **F7** site was removed by digestion with *Hinf* I this fragment did not bind GST-CREA. This suggests that like the situation 5' of *amdS*, one site is not sufficient for binding *in vitro*. A third region within the FACB1 fragment was protected, albeit to a lesser extent, in footprinting studies. The protected sequence does not contain any obvious GC rich sequences such as those found within the other protected regions of *facB* and *amdS*. Perhaps this represents a totally different recognition sequence for CREA that is dissimilar to the SYG GRG sequence found thus far. The question of another dissimilar binding site will be addressed in the following chapter with a degenerate oligonucleotide approach. The 130 bp *Nla* IV- *Taq* I fragment contains the second SYG GGG sequence, **F5**, but no other similar sequence. This fragment did not bind CREA strongly. To test whether a fragment containing the **F5** and **F6** sites but not the **F7** site, was able to bind CREA, a set of PCR primers were designed which amplify a region of 182 bp (shown in Fig. 3.12). There are 137 bp between these sites but perhaps two bound CREA molecules can interact by looping out the intervening DNA. Using these primers a fragment was obtained for use in a gel mobility shift assay (Fig. 3.16c) and found to exhibit some retardation but this was not as much

retardation as the larger 273 bp fragment. This suggests that perhaps the GST-CREA fusion protein requires that the two sites lie in close proximity for strong binding.

An analysis of the sequence 5' of *facB* shows that in addition to the two SYG GGG sequences (**F5** and **F6**) and the CTC CAG (**F7**) sequence, there are also four other SYG GAG sequences (**F1-F4**). Two large fragments containing these sequences (the 219 bp *Taq* I- *Nru* I and 239 bp *Sma* I - *Nhe* I fragments) were able to bind CREA but when digested, sub-fragments containing these sequences did not clearly bind CREA. This suggests that additional sequence elements required for binding were present in the larger fragments but not in the smaller fragments. This could be explained by the presence of a totally different recognition sequence for CREA from the SYG GRG sequence or perhaps some variation of this. Alternatively CREA may have a non-specific or general binding capacity for fragments of DNA of a certain size or composition. Two large fragments (ie. the 153 bp *Nru* I-*Sma* I and 157 bp *Taq* I-*Nhe* I fragments) contain SYG GAG sequences but did not bind, perhaps the surrounding context (5' and/ or 3') of these sites was not favourable for binding.

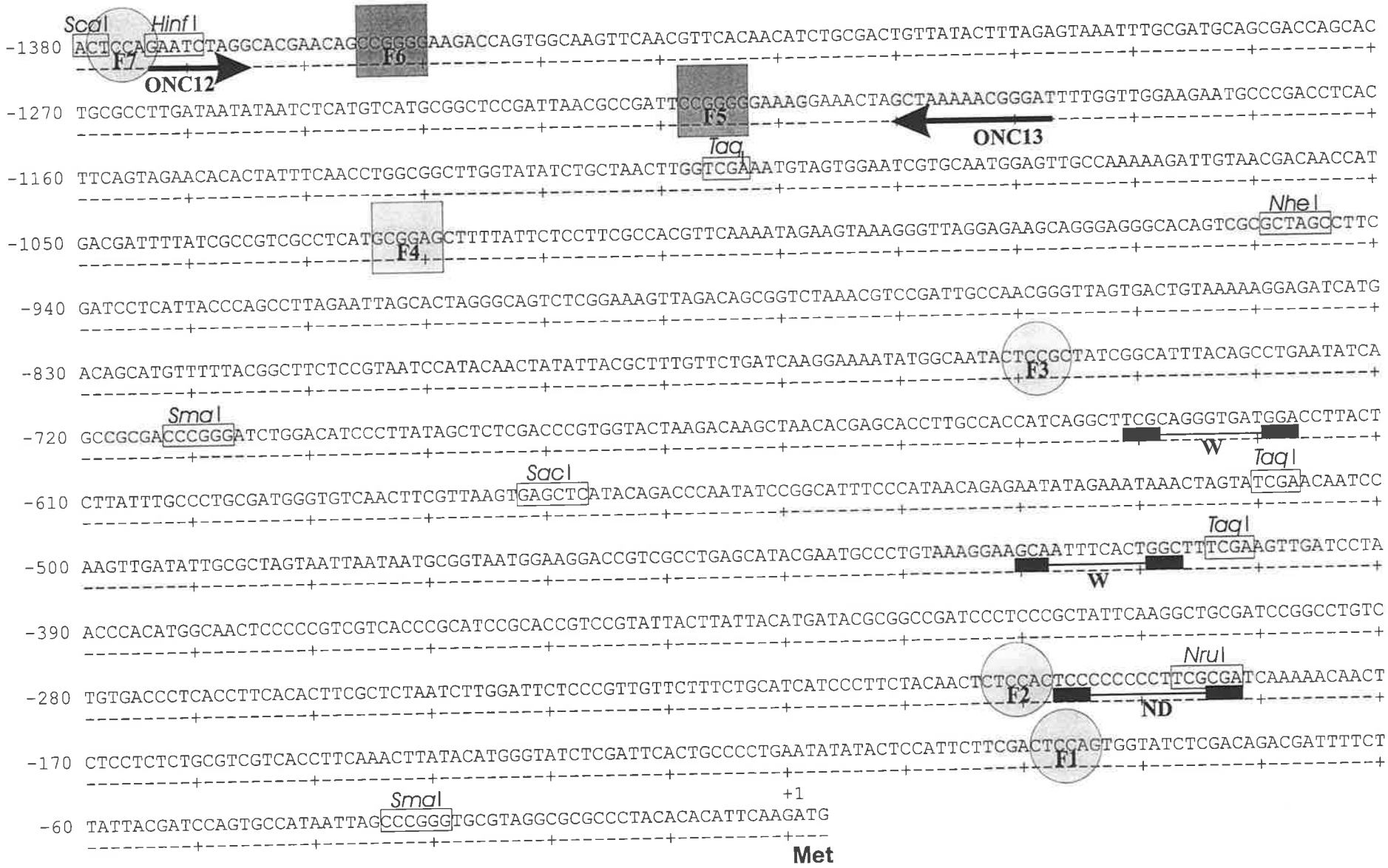
The two SYG GGG motifs (**F5** and **F6**) are located 1.21 and 1.35 kb upstream of the start point of translation and the CTC CAG site, **F7**, lies adjacent to **F6**. These sites are further upstream than the distance of **A1** from *amdS* which starts at -81. The start point of transcription for *facB* has not been determined (M.J. Hynes pers. comm.) and thus this may not be significant if *facB* has a long 5' untranslated region. At least two consensus SYG GGG sites are present in the 5' regions of the *facB* genes from *A. oryzae* and *A. niger* (Murphy, 1996). Evidence for the *in vivo* functionality of these two sites, **F5** and **F6**, requires mutation of the sites in order to observe the effects *in vivo*.

3.9.4 Recognition sites present in the 5' sequence of *creA*

The *creA204* mutant has higher levels of *creA* mRNA compared to wild type when grown in 1% glucose (Dowzer, 1991). This suggests that the negatively acting CREA protein may have an autoregulatory role. The finding that two fragments from the 5' region of *creA* bind GST-

Figure 3.22 Sequence of the 5' region of *facB*

Sequence of the the 5' region of the *facB* gene and some restriction endonuclease sites are shown (Todd *et al.*, submitted). All sequences fitting the consensus SYG GGG (dark shading) or the broader consensus SYG GRG (light shading) are shown and are labelled **F1** to **F7**. The orientation of the sites is shown by squares (G rich strand is the coding strand) or circles (G rich strand is the non-coding strand). The location of the two primers ONC12 and ONC13 is shown. The position of FACB binding sites (W= weak, ND= determined from sequence analysis only) is shown (Todd, 1995) and will be discussed in chapter 6.The start point of translation is +1.


ScaI
HinfI
F7
ONC12
F6
F5
ONC13
TaqI
NheI
F4
F3
SmaI
W
TaqI
SacI
TaqI
W
F2
NruI
ND
F1
SmaI
+1
Met
-1380 ACTCCAGAATCTAGGCACGAACAGCCGGGCAAGACCAGTGGCAAGTTCAACGTTCACAACATCTGCGACTGTTATACTTTAGAGTAAATTTGCGATGCAGCGACCAGCAC
-1270 TGCGCCTTGATAATATAATCTCATGTCATGCGGCTCCGATTAACGCCGATTCCGGGGGAAAGGAAACTAGCTAAAAACGGGATTTTGGTTGGAAGAATGCCCGACCTCAC
-1160 TTCAGTAGAACACACTATTTCAACCTGGCGGCTTGGTATATCTGCTAACTTGGTCGAAATGTAGTGGAATCGTGCAATGGAGTTGCCAAAAAGATTGTAACGACAACCAT
-1050 GACGATTTTATCGCCGTCGCCTCATGCGGAGCTTTTATTCTCCTTCGCCACGTTCAAAATAGAAGTAAAGGGTTAGGAGAAGCAGGGAGGGCACAGTCGCGCTAGCCTTC
-940 GATCCTCATTACCCAGCCTTAGAATTAGCACTAGGGCAGTCTCGGAAAGTTAGACAGCGGTCTAAACGTCCGATTGCCAACGGGTTAGTGACTGTAAAAAGGAGATCATG
-830 ACAGCATGTTTTTACGGCTTCTCCGTAATCCATACAACTATATTACGCTTTGTTCTGATCAAGGAAAATATGGCAATACTCCGCTATCGGCATTTACAGCCTGAATATCA
-720 GCCGCGACCCCGGGATCTGGACATCCCTTATAGCTCTCGACCCGTGGTACTAAGACAAGCTAACACGAGCACCTTGCCACCATCAGGCTTCGCAGGGTGATGGACCTTACT
-610 CTTATTTGCCCTGCGATGGGTGTCAACTTCGTTAAGTGAGCTCATACAGACCCAATATCCGGCATTTCCCATAACAGAGAATATAGAAATAAACTAGTATCGAACAATCC
-500 AAGTTGATATTGCGCTAGTAATTAATAATGCGGTAATGGAAGGACCGTCGCCTGAGCATACGAATGCCCTGTAAAGGAAGCAATTTCACTGGCTTTCGAAGTTGATCCTA
-390 ACCCACATGGCAACTCCCCCGTCGTCACCCGCATCCGCACCGTCCGTATTACTTATTACATGATACGCGGCCGATCCCTCCCGCTATTCAAGGCTGCGATCCGGCCTGTC
-280 TGTGACCCTCACCTTCACACTTCGCTCTAATCTTGGATTCTCCCGTTGTTCTTTCTGCATCATCCCTTCTACAACTCTCCACTCCCCCCCCTTCGCGATCAAAAACAACT
-170 CTCCTCTCTGCGTCGTCACCTTCAAACTTATACATGGGTATCTCGATTCACTGCCCCTGAATATATACTCCATTCTTCGACTCCAGTGGTATCTCGACAGACGATTTTCT
-60 TATTACGATCCAGTGCCATAATTAGCCCGGGTGCGTAGGCGCGCCCTACACACATTCAAGATG

CREA *in vitro* suggests that CREA may have a direct role in its autoregulation. An analysis of the sequence 5' of *creA* (Fig. 3.23) shows that there are three SYG GGG sites and three SYG GAG sites. The retarding fragments *Bam* HI-*Hinf* I and *Ava* I-*Hinf* I contain one of each of these sequences. The other two sites lie outside of the region tested with mobility shift analysis. Fragments without these sites did not bind GST-CREA, confirming the importance of these sites in binding by CREA. The **C5** site (5' of the *Bam* HI site) lies within a very GC rich region including sites that deviate from the SYG GRG consensus at only one position and as such may be functional. In particular there are 24 bp between **C5** and a sequence TTG GGG (on the non-coding strand). The **C3** and **C4** sites are not adjacent. In close proximity (20 bp) to the **C3** site is the non-consensus sequence TTG GGG (on the non-coding strand) and in close proximity (9 bp) to the **C4** site is the sequence GAG GGG (on the non-coding strand). Therefore it is possible that the reason this fragment binds CREA is because it contains two pairs of adjacent binding sites. The **C1** and **C2** sites are 12 bp apart, just over one turn of a β-DNA helix. The **C3** and **C4** sites are 5' of the start points of transcription reported by Dowzer and Kelly, (1991), whereas the **C1** and **C2** sites are situated between transcription start points found by 5' RACE analysis. Thus the **C3** and **C4** sites may regulate the expression of the larger transcripts and the **C1** and **C2** sites may be involved in regulating the smaller *creA* transcripts (Shroff *et al.*, submitted).

3.9.5 Comparisons with published data on GST-CREA binding

A number of researchers have used the same GST-CREA fusion protein that was used in this study to determine which sequences were recognised in the 5' region of other genes subject to carbon catabolite control.

Researchers working with *alcA*, *alcR*, the proline utilisation cluster (*prn*) and *ipnA* 5' regions have all used column purified protein extracts (Kulmburg *et al.*, 1993; Cubero and Scazzocchio, 1994; Espeso *et al.*, 1993; Espeso and Penalva, 1994). This may explain why they did not observe the endogenous nuclease activity for their DNase I footprinting experiments. However, under the conditions used in this study, the complexes with purified CREA protein did not

migrate into the gel, and specific retardation was not seen. Therefore total protein extracts were used. None of the researchers mentioned that they treated the extract before column purification. These researchers did however, use a different *E. coli* strain to produce their fusion protein and perhaps this explains the different results with the column purified extract. These researchers report a yield of 1-2 μg/ μl of fusion protein (Kulmburg *et al.*, 1993; Cubero and Scazzocchio, 1994) which is similar to the 0.8-1.6 μg/ μl that was obtained for use in this study.

3.9.5.1 Alcohol metabolism

As discussed in section 1.3.3 there is evidence that *alcA*, *alcR* and *aldA* are under direct control of the *creA* gene product. Gel mobility shift and DNase I sensitivity assays on the 5' region of the *alcR* gene has determined four binding sites for CREA which fit the MIG1 consensus SYG GGG (Fig. 3.24, Kulmburg *et al.*, 1993), consistent with the results obtained in this study. The **A** and **B** sites are located upstream of the start point of transcription whereas the **C1** and **C2** sites are not. An analysis of the available 5' sequence shows that five SYG GAG sequences are also present. One of these, **A'**, lies 8 bp down stream of the **A** site and was also protected by CREA but the other, **C'**, was reported to be protected over only half the sequence (Fig. 3.24). The situation 5' of *alcR* appears similar to what was found in this study for *amdS*, *creA* and to a lesser extent *facB* where multiple binding sites were present, some of which were adjacent to each other. Binding sites for the positive autoregulator ALCR 5' of *alcR* have been determined using an *E. coli* expressed GST-ALCR fusion protein (Kulmburg *et al.*, 1992a). Two pairs of sites were found, one a direct and the other an inverted repeat of the sequence 5' CCGCA 3' . The GST-ALCR binding sites overlap with the **A** and **A'** CREA binding sites (Kulmburg *et al.*, 1992a, 1993). This suggests that repression by CREA may involve competition between the positive regulator ALCR and the negatively acting CREA protein for binding to sites 5' of *alcR*. Although there are many binding sites for CREA within the *alcR* promoter, Kulmburg *et al.*, (1993) postulate that the **A**, and presumably the **A'**, sites are the major functional CREA binding sites since they overlap with the GST-ALCR binding sites. To test the *in vivo* functionality of these sites, these researchers made constructs containing large deletions of the

Figure 3.23 Sequence of the 5' region of *creA*

Sequence of the the 5' region of the *creA* gene and some restriction endonuclease sites is shown (Dowzer and Kelly, 1991; Shroff *et al.*, submitted). All sequences fitting the consensus SYG GGG (dark shading) or the broader consensus SYG GRG (light shading) are shown and are labelled **C1** to **C6**. The orientation of the sites is shown by circles (G rich strand is the non-coding strand). The position of start points of transcription reported in Dowzer and Kelly (1991) are shown with #. Those reported in Shroff *et al.*, (submitted) are shown with an "o". Potential binding sites for AREA are underlined and a potential site for the PACC protein is boxed (Merika and Orkin, 1993; Tilburn *et al.*, 1995). The start point of translation is +1.

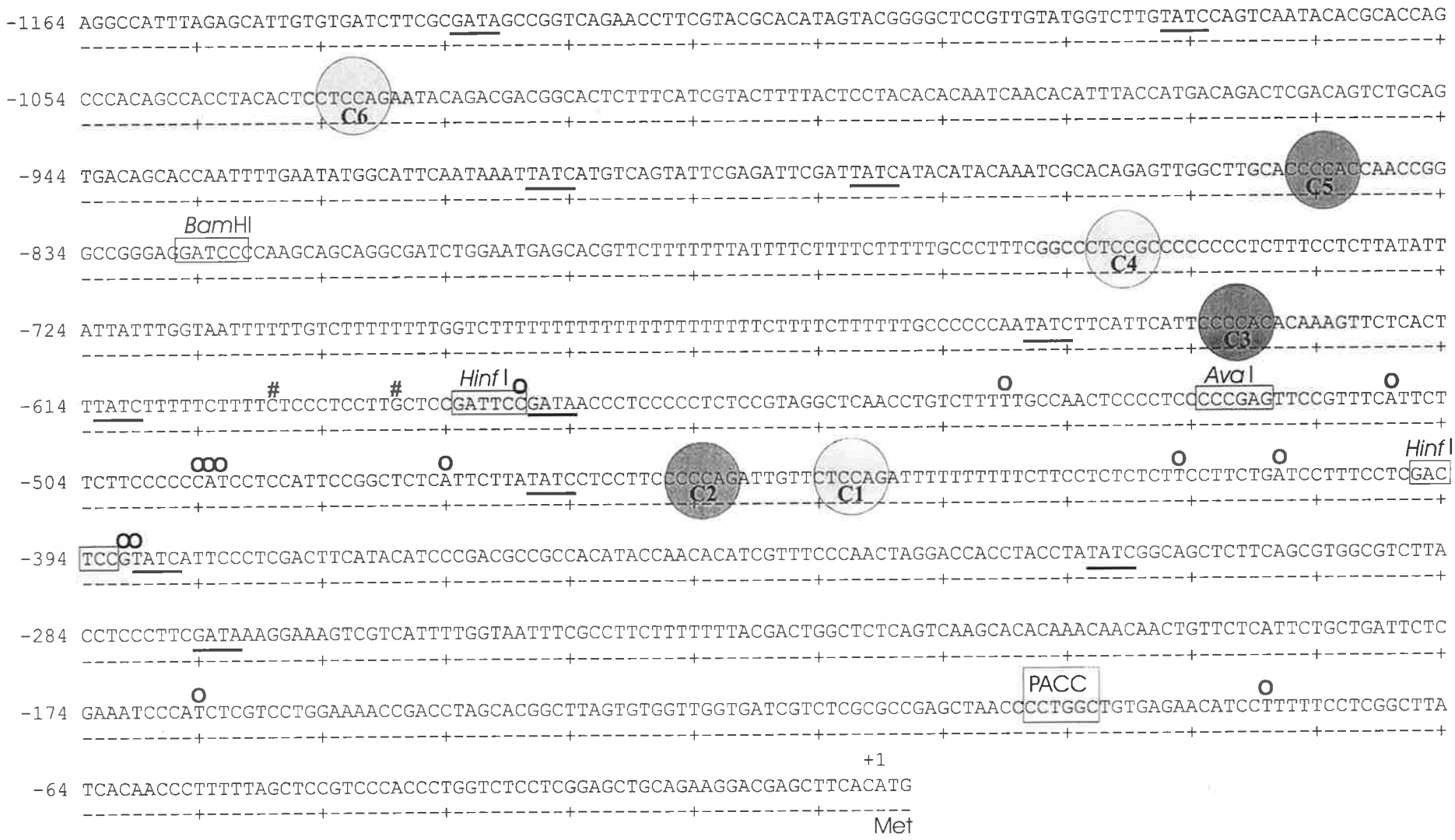
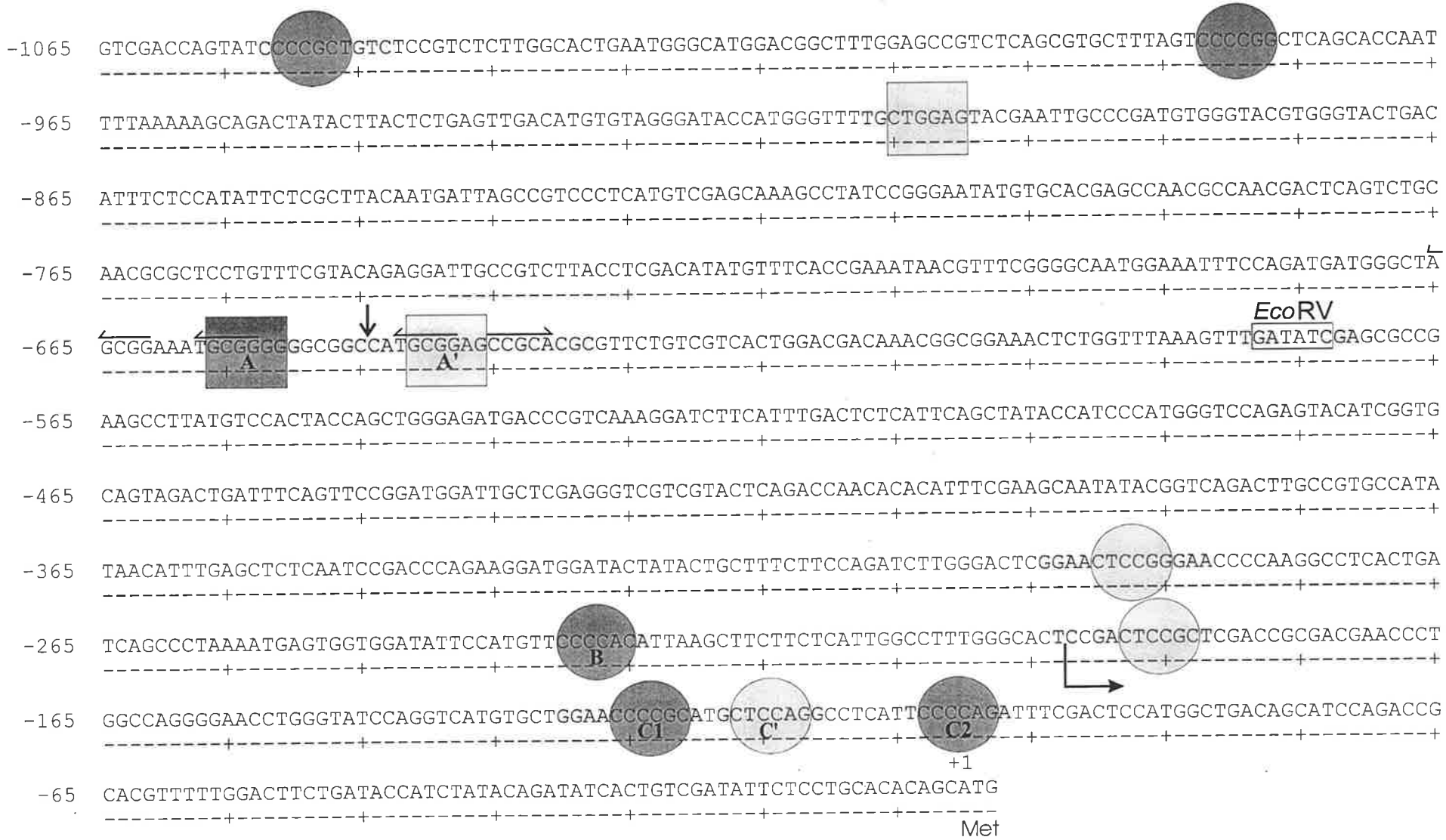

-1164 AGGCCATTTAGAGCATTGTGTGATCTTCGCGATAGCCGGTCAGAACCTTCGTACGCACATAGTACGGGGCTCCGTTGTATGGTCTTGTATCCAGTCAATACACGCACCAG
-1054 CCCACAGCCACCTACACTCCTCCAGAATACAGACGACGGCACTCTTTCATCGTACTTTTACTCCTACACACAATCAACACATTTACCATGACAGACTCGACAGTCTGCAG
C6
-944 TGACAGCACCAATTTTGAATATGGCATTCAATAAATTATCATGTCAGTATTCGAGATTCGATTATCATACATACAAATCGCACAGAGTTGGCTTGCACCCCACCAACCGG
C5
BamHI
-834 GCCGGGAGGGATCCCCAAGCAGCAGGCGATCTGGAATGAGCACGTTCTTTTTTTATTTTCTTTTCTTTTTGCCCTTTCGGCCCTCCGCCCCCCCCTCTTTCCTCTTATATT
C4
-724 ATTATTTGGTAATTTTTTGTCTTTTTTTTGGTCTTTTTTTTTTTTTTTTTTTTTTCTTTTCTTTTTTGCCCCCCAATATCTTCATTCATTCCCCACACAAAGTTCTCACT
C3
Hinf I
Ava I
-614 TTATCTTTTTCTTTTCTCCCTCCTTGCTCCGATTCCGATAACCCTCCCCCTCTCCGTAGGCTCAACCTGTCTTTTTGCCAACTCCCCTCCCCCGAGTTCCGTTTCATTCT
Hinf I
-504 TCTTCCCCCCCATCCTCCATTCCGGCTCTCATTCTTATATCCTCCTTCCCCCAGATTGTTCTCCAGATTTTTTTTTTCTTCCTCTCTCTTCCTTCTGATCCTTTCCTCGAC
C2
C1
-394 TCCGTATCATTCCCTCGACTTCATACATCCCGACGCCGCCACATACCAACACATCGTTTCCCAACTAGGACCACCTACCTATATCGGCAGCTCTTCAGCGTGGCGTCTTA
-284 CCTCCCTTCGATAAAGGAAAGTCGTCATTTTGGTAATTTCGCCTTCTTTTTTTACGACTGGCTCTCAGTCAAGCACACAAACAACAACTGTTCTCATTCTGCTGATTCTC
PACC
-174 GAAATCCCATCTCGTCCTGGAAAACCGACCTAGCACGGCTTAGTGTGGTTGGTGATCGTCTCGCGCCGAGCTAACCCCTGGCTGTGAGAACATCCTTTTTCCTCGGCTTA
+1
-64 TCACAACCCTTTTTAGCTCCGTCCCACCCTGGTCTCCTCGGAGCTGCAGAAGGACGAGCTTCACATG
Met

Figure 3.24 GST-CREA binding sites present in the 5' region of the *alcR* gene

The sequence of the 5' region of *alcR* is presented (Felenbok *et al.*, 1988). The amended start points of translation and transcription (-187) are indicated (Kulmburg *et al.*, 1991; Felenbok *et al.*, 1988). The position of the GST-CREA binding sites determined by DNase I sensitivity assays and the additional site **A'** (Mathieu and Felenbok, 1994) are shown (Kulmburg *et al.*, 1993). The additional site **C'** discussed in the text is also shown. All sequences fitting the consensus SYG GGG (dark shading) or the broader consensus SYG GRG (light shading) are shown, including those not reported to be protected in DNase I sensitivity assays (unlabelled sites). The orientation of the sites is shown by squares (G rich strand is the coding strand) or circles (G rich strand is the non-coding strand). The position of the binding sites for the GST-ALCR fusion protein determined by DNase I and methylation protection studies (Kulmburg *et al.*, 1992a) are shown with single headed arrows.

The deletion resulting in deletion of the ALCR binding sites starts at the *Eco* RV site shown and continues until -1064 (Kulmburg *et al.*, 1992a). The deletion used to create *alcRT*ΔA starts between the **A** and **A'** CREA binding sites at position -645 and continues for 419 bp until -1064 (Kulmburg *et al.*, 1993). The phenotype of strains containing the above two deletions will be discussed in chapter 6.


-1065 GTCGACCAGTATCCCCGCTGTCTCCGTCTCTTGGCACTGAATGGGCATGGACGGCTTTGGAGCCGTCTCAGCGTGCTTTAGTCCCCGGCTCAGCACCAAT
-965 TTTAAAAAGCAGACTATACTTACTCTGAGTTGACATGTGTAGGGATACCATGGGTTTTGCTGGAGTACGAATTGCCCGATGTGGGTACGTGGGTACTGAC
-865 ATTTCTCCATATTCTCGCTTACAATGATTAGCCGTCCCTCATGTCGAGCAAAGCCTATCCGGGAATATGTGCACGAGCCAACGCCAACGACTCAGTCTGC
-765 AACGCGCTCCTGTTTCGTACAGAGGATTGCCGTCTTACCTCGACATATGTTTCACCGAAATAACGTTTCGGGGCAATGGAAATTTCCAGATGATGGGCTA
EcoRV
-665 GCGGAAATCCGGCGGGGCGGCCATGCGGAGCCGCACGCGTTCTGTCGTCACTGGACGACAAACGGCGGAAACTCTGGTTTAAAGTTTGATATCGAGCGCCG
A
A'
-565 AAGCCTTATGTCCACTACCAGCTGGGAGATGACCCGTCAAAGGATCTTCATTTGACTCTCATTCAGCTATACCATCCCATGGGTCCAGAGTACATCGGTG
-465 CAGTAGACTGATTTCAGTTCCGGATGGATTGCTCGAGGGTCGTCGTACTCAGACCAACACACATTTCGAAGCAATATACGGTCAGACTTGCCGTGCCATA
-365 TAACATTTGAGCTCTCAATCCGACCCAGAAGGATGGATACTATACTGCTTTCTTCCAGATCTTGGGACTCGGAACTCCGGGAACCCCAAGGCCTCACTGA
-265 TCAGCCCTAAAATGAGTGGTGGATATTCCATGTTCCCCACATTAAGCTTCTTCTCATTGGCCTTTGGGCACTCCGACTCCGCTCGACCGCGACGAACCCT
B
-165 GGCCAGGGGAACCTGGGTATCCAGGTCATGTGCTGGAACCCCGCATGCTCCAGGCCTCATTCCCCAGATTTCGACTCCATGGCTGACAGCATCCAGACCG
C1
C'
C2
+1
-65 CACGTTTTTGGACTTCTGATACCATCTATACAGATATCACTGTCGATATTCTCCTGCACACAGCATG
Met

alcR promoter and these will be discussed in chapter 6. Footprinting and the other evidence discussed in chapter 6 suggests that the **A** and **A'** CREA binding sites are likely to be active *in vivo* and that the other sites **B**, **C1** and **C2** may contribute to repression.

Binding to the 5' region of *alcA* by CREA was also found and is shown in Figure 3.25a (Kulmburg *et al.*, 1993). Two protected regions were seen, **A** and **B**, and both of these contain a SYG GGG sequence (Fig. 3.25a). An analysis of the 5' *alcA* sequence for the broader consensus SYG GAG shows that at least two such sequences are present, one in close proximity to the **A** CREA binding site and the other in close proximity to the **B** binding site. Kulmburg *et al.*, (1993), do not report that these sites are protected in footprinting experiments. In addition to these there is a sequence adjacent to the **B** site which differs from the SYG GRG consensus at only one position. Both **A** and **B** pairs of sites are 5' to the start point of transcription. Binding sites for the ALCR protein 5' of *alcA* have been determined and were found to be 6 bp upstream of the **A** CREA binding site and 18 bp upstream of the **B** site (Kulmburg *et al.*, 1992b). Once again competition between ALCR and CREA may be a feature of the mechanism of repression (Kulmburg *et al.*, 1993). The **A** and **B** sites from *alcA* have been mutated *in vitro* and this construct used to transform a strain expressing ALCR under the constitutive *gpdA* promoter (Hintz and Lagosky, 1993). Expression of ALCA was found to be derepressed in the presence of the glucose analogue 2-deoxyglucose, suggesting that these sites are functional *in vivo*. A strain producing ALCR constitutively under the *gpdA* promoter is only partially repressed for *alcA* but totally repressed for *aldA* at the level of transcription (Kulmburg, 1991, cited by Felenbok and Sealy-Lewis, 1994). This suggests that ALCR is able to compete with CREA 5' of *alcA* resulting in only partial repression of *alcA* by CREA. In contrast increased levels of ALCR do not result in derepression at the *aldA* locus and therefore repression by CREA cannot involve competition with ALCR protein. No GST-ALCR binding sites 5' of *aldA* were found, either by sequence analysis or by mobility shift assays (Pickett *et al.*, 1987; Kulmburg, 1991, cited by Felenbok and Sealy-Lewis, 1994). This may be because *aldA* is not directly activated by ALCR. No binding studies of CREA to the 5' region of *aldA* have been reported in the

literature, however a number of CREA binding sites can be found and these will be discussed in chapter 6.

The protected regions of these genes are all GC rich and an alignment of the SYG GRG sites (Fig. 3.25b) shows that all the *alcR* sites, but none of the *alcA* sites, contain a T in the seventh position. In addition the *alcR* sites show a preference for A's and T's in positions 8 to 11 (immediately 5' of the core sequence) but this is not observed for the *alcA* sites. Although a double block mechanism of repression acts on both the regulatory (*alcR*) and the structural gene (*alcA*) perhaps the CREA binding sites 5' of *alcR* are stronger since they approximate the MIG1 binding site more closely. It may have followed that repression of the *alcA* locus became less dependent on the CREA sites 5' of *alcA* and thus these sites diverged to become weaker. An alignment of the CREA sites 5' of *amdS* and *facB* (Errata, Figure 1) shows little about the preferences 5' and 3' and this is probably due to the close proximity of sites.

3.9.5.2 The proline utilisation cluster

As discussed in section 1.3.4 the proline utilisation cluster is subject to carbon catabolite repression. Cubero and Scazzocchio (1994) have used mobility shift analysis, DNase I and methylation protection studies to located all of the sites within the 1.6-1.7 kb *prnD-prnB* intergenic region which were able to bind the GST-CREA fusion protein *in vitro*. These researchers report binding to seven sites. The sequence 699 bp upstream of the *prnB* gene is available (Sophianopoulou and Scazzocchio, 1989) and contains five of these sites (Fig. 3.26). Sites **1**, **2**, **3.1**, **4.1** and **4.2** conform to the MIG1 consensus, SYG GGG. One site, **3.2**, conforms to the broader consensus SYG GAG. The site **3.3** deviates from SYG GRG at one position but was still protected in DNase I, methylation protection and depurination interference studies (Cubero and Scazzocchio, 1994). Perhaps its location in close proximity to the sites **3.1** and **3.2** allowed this site to be bound by GST-CREA. As was found 5' of *alcR* and *alcA* multiple GST-CREA binding sites, some of which are adjacent, were found within the *prn* intergenic region. Cloning and sequencing of the three *cis*-acting mutations which alleviate carbon catabolite

Figure 3.25a GST-CREA binding sites present in the 5' region of the *alcA* gene

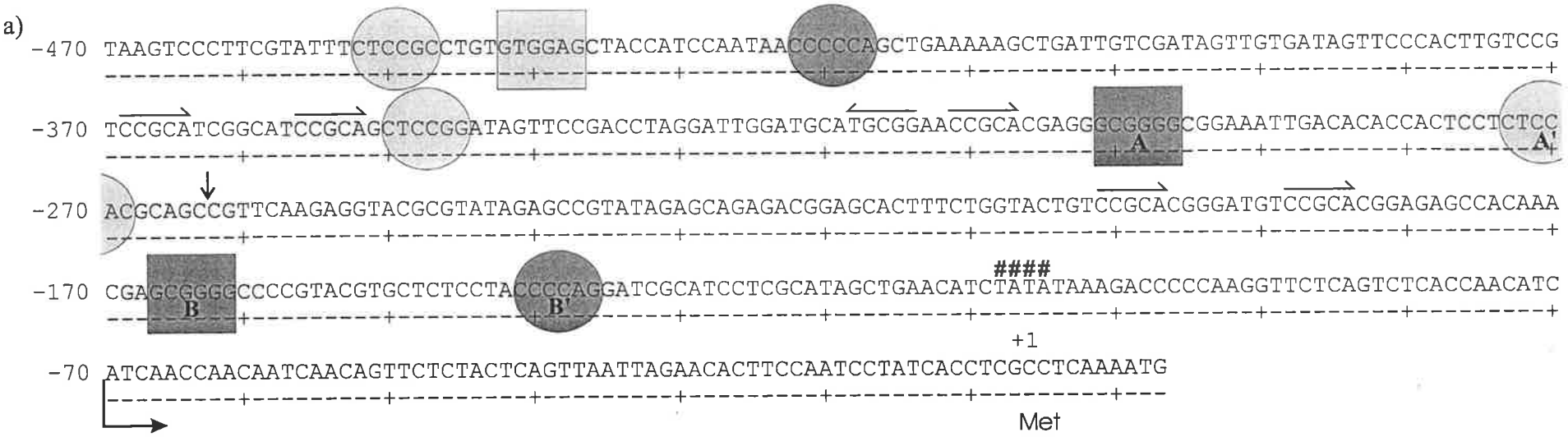
The sequence of the 5' region of *alcA* is presented (Gwynne *et al.*, 1987; Kulmburg *et al.*, 1992b). The start points of translation and transcription (-69) are indicated (Gwynne *et al.*, 1987). The position of the GST-CREA binding sites determined by DNase I sensitivity assays are shown (Kulmburg *et al.*, 1993) as are the additional sites **A'** and **B'** discussed in the text. All sequences fitting the consensus SYG GGG (dark shading) or the broader consensus SYG GRG (light shading) are shown, including those not reported to be protected in DNase I sensitivity assays (unlabelled sites and sites **A'** and **B'**). The orientation of the sites is shown by squares (G rich strand is the coding strand) or circles (G rich strand is the non-coding strand).The position of the binding sites for the GST-ALCR fusion protein determined by DNase I and methylation protection studies (Kulmburg *et al.*, 1992a) are shown with single headed arrows.

The deletion resulting in deletion of two thirds of the ALCR binding sites and the two CREA binding sites **A** and **A'** ends at the position marked with the vertical arrow (Kulmburg *et al.*, 1992b). The phenotype of a strain containing the above deletion will be discussed in chapter 6.

Figure 3.25b Alignment of GST-CREA binding sites from the *alcR* and *alcA* genes

An alignment of the sites fully protected by GST-CREA in DNase I sensitivity assays of the *alcR* and *alcA* 5' regions (Kulmburg *et al.*, 1993) is shown.

a)



b)

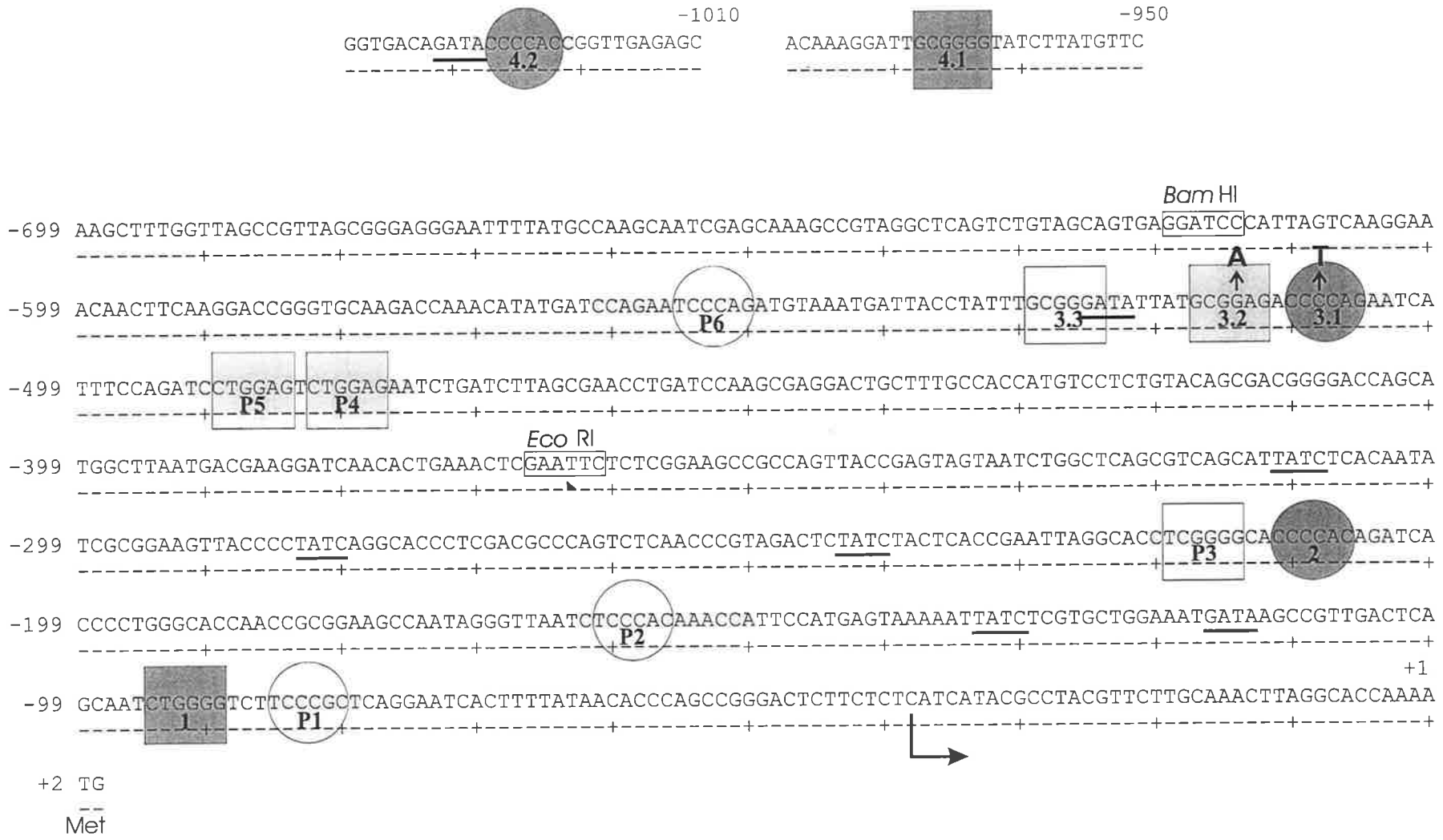
alcR	CG GAAA T GCG GGG GGCGGCCAT CG GCCA T GCG GAG CCGCACGCG GC TTAA T GTG GGG AACATGGAA GG AGCA T GCG GGG TTCCAGCAC TC GAAA T CTG GGG AATGAGGCC
consensus	**AA T GCG GGG** **C CT A**

alcA	CA CGAG G GCG GGG CGGAAATTGA CA AACG A GCG GGG CCCCGTACGT
consensus	**G GCG GGG** **A**

Figure 3.26 GST-CREA binding sites present in the intergenic region of the *prn* cluster

The sequence of the *prnD-prnB* intergenic region up to -699 bp (from *prnB*) and the region surrounding the binding sites **4.1** and **4.2** is presented (Sophianopoulou and Scazzocchio, 1989; Cubero and Scazzocchio, 1994). The start points of translation and transcription (-38) for the *prnB* gene are indicated (Sophianopoulou and Scazzocchio, 1989). The position of the GST-CREA binding sites determined by DNase I sensitivity assays and methylation protection studies are shown (Cubero and Scazzocchio, 1994). Sequences fitting the consensus SYG GGG (dark shading), the broader consensus SYG GRG (light shading) or the sequence SYG GGA (unshaded) are shown. The orientation of the sites is shown by squares (G rich strand is the coding strand) or circles (G rich strand is the non-coding strand). The additional sites **P1- P6** are discussed in the text. The position of potential AREA binding sites (GATA/ TATC sequences; Merika and Orkin, 1993) are underlined for sequences on both strands. The PRNA binding sites (not shown) have been reported not to be in close proximity to those of CREA (Cazelle, 1993, cited by Scazzocchio *et al.*, 1995).

The *prn^{d}20* mutation is a G3-->A3 transition within the **3.2** site and the *prn^{d}21* and *prn^{d}22* mutations are both C3-->T3 transitions within the **3.1** site (Sophianopoulou *et al.*, 1993).


-1010
GGTGACAGATACCCCACCGGTTGAGAGC
4.2
-950
ACAAAGGATTGCGGGGTATCTTATGTTC
4.1
Bam HI
-699 AAGCTTTGGTTAGCCGTTAGCGGGAGGGAATTTTATGCCAAGCAATCGAGCAAAGCCGTAGGCTCAGTCTGTAGCAGTGAGGATCCCATTAGTCAAGGAA
A
T
-599 ACAACTTCAAGGACCGGGTGCAAGACCAAACATATGATCCAGAATCCCAGATGTAAATGATTACCTATTTGCGGGATATTATGCGGAGACCCCAGAATCA
P6
3.3
3.2
3.1
-499 TTTCCAGATCCTGGAGTCTGGAGAATCTGATCTTAGCGAACCTGATCCAAGCGAGGACTGCTTTGCCACCATGTCCTCTGTACAGCGACGGGGACCAGCA
P5
P4
Eco RI
-399 TGGCTTAATGACGAAGGATCAACACTGAAACTCGAATTCTCTCGGAAGCCGCCAGTTACCGAGTAGTAATCTGGCTCAGCGTCAGCATTATCTCACAATA
-299 TCGCGGAAGTTACCCCTATCAGGCACCCTCGACGCCCAGTCTCAACCCGTAGACTCTATCTACTCACCGAATTAGGCACCTCGGGGCACCCCACAGATCA
P3
2
-199 CCCCTGGGCACCAACCGCGGAAGCCAATAGGGTTAATCTCCCACAAACCATTCCATGAGTAAAAATTATCTCGTGCTGGAAATGATAAGCCGTTGACTCA
P2
+1
-99 GCAATCTGGGGTCTTCCCGCTCAGGAATCACTTTTATAACACCCAGCCGGGACTCTTCTCTCATCATACGCCTACGTTCTTGCAAACTTAGGCACCAAAA
1
P1
+2 TG
Met

repression of the *prn* cluster has revealed that they are mutations within the pair of adjacent sites **3.1** and **3.2** (Sophianopoulou *et al.*, 1993). The *prn^{d}20* mutation is a G3-->A3 transition within a T GCG GAG site and the *prn^{d}21* and *prn^{d}22* mutations are both C3-->T3 transitions within a T CTG GGG site (non-coding strand) (see Fig. 3.26). Binding by the GST-CREA fusion protein to the **3.1** and **3.2** sites was abolished in the presence of the *prn^{d}22* and *prn^{d}20* mutations respectively and thus the *in vivo* functionality of these sites was shown to be a direct result of interactions with the CREA protein. In addition to the five sites protected by CREA, two additional sites are present, **P4** and **P5**, which conform to the broader sequence SYG GRG, as did the site **3.2**, but **P4** and **P5** were not reported to be protected. The protected site **3.3** deviates from the SYG GGG consensus at position 1. one site, **P1**, is present which has the same core sequence as **3.3**, GCG GGA sequence, albeit on the non-coding strand (Fig. 3.26). Furthermore **P1**, is adjacent to a site, **1**, which does bind CREA but site **P1** was not reported to be protected in DNase I sensitivity assays. Two other sites, **P2** and **P6**, are present which fit the sequence, T SYG GGA (non-coding strand; Fig. 3.26) and neither of these extra sites were found to be protected. One site, **P3**, was reported to be protected in some DNase I sensitivity assays.

These researchers made oligonucleotides containing the **3.2** binding site and the non-functional *prn^{d}22* site. The sequence of these oligonucleotides, O1, O2 and O3 are presented below:

O1 GATATTATGCGGAGACCTCAGA

O2 GATATTATG**T**GGAGACCTCAGA

O3 GATATTAT**C**CGGAGACCTCAGA

These oligonucleotides all contain a perfect AT rich region and a T7 immediately 5' of the core GC binding sequence. They differ only in the nucleotide at position 5 or 6. The first oligonucleotide O1 contains a single weak site and was found to bind GST-CREA with only a very weak affinity. The affinity of an oligonucleotide is scored in this study based on the percentage of

bound oligonucleotide to unbound oligonucleotide. In the case of O1 much less than 50% was bound and therefore it is said to bind very weakly. Oligonucleotides from the 5' region of *amdS* which contained a single site were also shown to bind GST-CREA with only a very weak affinity. Oligonucleotide O3 was found not to bind GST-CREA at all and this suggests that changing the core sequence can alter the affinity of the oligonucleotide to below the level at which it will bind. When O2 was tested it was found to have average binding *in vitro* to the GST-CREA fusion protein, since approximately 50% of this oligonucleotide was bound. This suggests that oligonucleotides containing a single site can indeed bind GST-CREA *in vitro* and is in contrast to results obtained with the oligonucleotides from the 5' region of *amdS* (section 3.8.2). Perhaps the presence of a perfect AT rich sequence and a T7 5' of the core GC sequence allows oligonucleotides with only one site to bind with average affinity. These researchers used much less GST-CREA protein than was used in this study and therefore no direct comparisons between the binding affinity of oligonucleotides can be made. However, oligonucleotide ONC1/ONC17 from the 5' *amdS* region which contains three potential sites appeared to bind with higher affinity than O2 therefore perhaps higher affinity binding results from additional sites and that O2 is the bare minimum for average but not strong binding. The context of a potential CREA binding site appears very important for binding.

These studies confirm that CREA recognises an SYG GGG binding site but expand this consensus to at least include SYG GAG. Although only a few genes have been studied it seems likely that adjacent sites are important in controlling genes regulated by CREA *in vivo*. This is due to the fact that mutations in one or other of the CREA binding sites within the *prn* intergenic region results in derepression. This was also the case *in vivo* for the CREA sites, **A1** and **A2**, within the ALU148 protected region of *amdS* and for binding *in vitro* for the **F6** and **F7** sites 5' of *facB*. The presence of more than one site clearly improves binding by CREA but results with the oligonucleotide O2 show that a single site with the correct context is able to bind GST-CREA *in vitro*. To what extent, if any, the other many sites present in the *prn*

intergenic region contribute to repression is not known. Perhaps additional sites serve to increase the local concentration of CREA protein.

3.9.5.3 Penicillin biosynthesis

The secondary metabolism gene encoding isopenicillin N synthase, (*ipnA*), plays a key role in the biosynthesis of penicillins and cephalosporins. The *ipnA* gene is under carbon catabolite repression (Espeso and Penalva, 1992). Mutations in *creA* only slightly derepress *ipnA* synthesis in the presence of sucrose, and this could be due to non-heirachical heterogeneity or that CREA does not have a role in carbon catabolite repression of this gene (Espeso and Penalva, 1992).

Perez-Esteban *et al.*, (1993) have shown that the *ipnA* gene is divergently transcribed from the ACV-synthase gene of the same cluster, with a distance of 526 bp between the two start points of transcription and a negative carbon controlling element for *ipnA* lies within the coding region of the ACV-synthase gene. Espeso and Penalva (1994) have found multiple GST-CREA binding sites within the 5' upstream region of *ipnA*. Espeso *et al.*, (1993) have shown that the only *in vitro* protected CREA binding site within this negative controlling region was not active *in vivo* as deletion of this site has no effect on the expression of a 5' *ipnA-lacZ* fusion construct. It seems unlikely that CREA has a direct or even an indirect role in repression of *ipnA* expression. Results presented in this study show that CREA did not bind to fragments containing sites divergent from the MIG1 consensus. Conditions under which binding is carried out, for example salt or protein concentrations, will affect the strength of the binding reaction, and in particular may affect the number of contacts required between CREA and the DNA helix for strong binding. Differences such as these may explain why these researchers showed binding by GST-CREA to many sites within the 5' *ipnA* region many of which diverge significantly from the MIG1 consensus and why these are unlikely to be active *in vivo* (Espeso and Penalva, 1994). These researchers report that sites which diverged significantly from the consensus were always located in close proximity to sites which did not.

3.9.6 Concluding remarks

There are many CREA binding sites in the 5' upstream regions of all the genes studied thus far and it seems likely that more than one site is required for repression by CREA. The presence of CREA sites in close proximity to binding sites for other regulatory proteins suggests that competition for efficient binding may be a feature of co-ordinated regulation. It appears that CREA recognises a similar binding site to MIG1 but that *in vivo* divergent sites like the one defined by the *prn^{d}20* mutation are functional. Only one functional divergent site has been found so far for MIG1. Like MIG1, CREA binding appears to be very context depended however unlike MIG1 whose sites found thus far do not differ greatly, CREA appears to recognise a broader range of binding sites and may prefer two sites within the same DNA fragment for binding. An AT rich region immediately 5' of the recognition sequence was found in all naturally occurring MIG1 sites. To determine whether CREA binding sites contain a 5' or 3' extended context it is possible to take a natural binding site and mutate every base. However, an easier approach is to use degenerate oligonucleotides to define the recognition sequence further. Using this approach it is theoretically possible to find out what bases are required, whether there is an extended context and whether binding really does require two recognition sites and if so what spacing between them is acceptable and whether orientation is important. The results of such an approach is presented in the following chapter. Another feature of this approach is that it will allow the identification of a totally novel binding site if one exits.

CHAPTER 4

Chapter 4

Target Selection by GST-CREA From Random Sequences

In this chapter experiments are reported in which the GST-CREA fusion protein was used to select oligonucleotides from a pool of random oligonucleotides.

The aims of this experiment were:

1. to determine which sequences the GST-CREA fusion protein can recognise
2. to determine whether there is a preference for particular nucleotides 5' and 3' of the consensus sequence so determined.

4.1 CASTing (Cyclic amplification and selection of targets)

CASTing (reviewed by Wright and Funk, 1993) is a technique which can be used to determine the recognition sequence(s) of a DNA binding protein when the *in vitro* conditions required for binding are known. CASTing requires the synthesis of oligonucleotides which contain a degenerate region (consisting of a fixed number of nucleotides of random sequence), flanked by defined sequences that allow PCR amplification by primers. Oligonucleotides containing the recognition site can be selected from this pool of oligonucleotides by a number of methods including immunoprecipitation, affinity column purification and mobility shift analysis. A number of cycles of selection and amplification are carried out (commonly three) before oligonucleotides are cloned and sequenced. Sequences can be analysed for nucleotide strings common to all oligonucleotides. In this way a number of researchers have been able to obtain more information on the recognition sequences of particular DNA binding proteins (Perkins *et al.*, 1991; Merika and Orkin, 1993). In particular, a significant variation of the known binding site for the retinoblastoma containing complex was found using CASTing (Ouellette *et al.*, 1992).

4.1.1 Generation of a pool of oligonucleotides for CASTing

A pool of oligonucleotides 687 (Table 2.3) was synthesised which contained a 20 bp variable region flanked by 10 bp of defined sequence, the 5' end identical to10 bp of oligonucleotide 602 and the 3' end complementary to 10 bp of oligonucleotide 603.

To check that the degenerate oligonucleotides (687) synthesised were indeed truly random, they were made double stranded and cloned into pGEM-T. Ten clones were selected at random and sequenced. From the sequences of these unselected oligonucleotides (Table 4.1) no obvious preference for a particular nucleotide at any position was found, and there was no imbalance in the ratio of A:C:G:T. The clone 01 contained a variable region of only 19 bp. This was most likely due to a lack of incorporation of a base during the synthesis of the 687 oligonucleotides perhaps due to failure at the deblocking or coupling step.

4.1.2 Selection and cloning of oligonucleotides containing the CREA binding site

To select oligonucleotides that contained the CREA binding site, oligonucleotide pool 687 was made double stranded, in the presence of radiolabelled dATP, using 2μg of oligonucleotide 603 and one round of annealing and extension with Taq polymerase (section 2.2.8.1). The double stranded oligonucleotides were ethanol precipitated and resuspended in 12μl of PBS, and used in a mobility shift assay with the *E. coli* expressed GST-CREA fusion protein. In initial experiments a glutathione sepharose column was used to purify the oligonucleotides bound to GST-CREA, however, after sequencing these products, it was clear that large multimeric PCR products had been formed via primer dimerisation. This was overcome by the use of a gel mobility shift assay to purify oligonucleotides bound to GST-CREA. A mobility shift assay was carried with the double stranded oligonucleotide 687 pool and in order to visualise the position of free (unretarded) oligonucleotides, EDTA was added to a final concentration of 200 mM to 1/30th of the mixture and the samples electrophoresed. After autoradiography (Fig. 4.1a) the low mobility (retarded) complexes were excised (indicated by box A), electrophoresed onto DEAE membrane and eluted. The selected oligonucleotides were amplified by the PCR conditions outlined in section 2.2.8.2

After this first round of selection and amplification, oligonucleotides were subject to another round of selection as described above except that the unretarded (free) oligonucleotides were also excised in order to determine whether the unretarded oligonucleotides were still capable

Table 4.1 Sequence of ten unselected oligonucleotides from the 687 pool

Name	Sequence[a]
01	AATTGAACGC**ACATGTCTGCACGAGGGCG.**GGTTACACCATTGA
B5	AATTGAACGC**GGTGCAGCACGGGCATAACG**GGTTACACCATTGA
03	AATTGAACGC**TGTTCCTGTTCACCACGCGA**GGTTACACCATTGA
04	AATTGAACGC**TTCCGAATCACTGCATTAAG**GGTTACACCATTGA
05	AATTGAACGC**TTATGCGGTCGGCTCAAGTG**GGTTACACCATTGA
06	AATTGAACGC**AGACACCCGTCCCGGATGCC**GGTTACACCATTGA
07	AATTGAACGC**TTGCAGCGACGCAGCCGTCA**GGTTACACCATTGA
08	AATTGAACGC**GCTGTTCTTGAAGATCTCGC**GGTTACACCATTGA
10	AATTGAACGC**AATGCACCTGGATACTAGGA**GGTTACACCATTGA
11	AATTGAACGC**GTCTGGCTATGGGGGGTGCG**GGTTACACCATTGA

[a] base composition:

		total
A	31302310312313323213	40
C	02233363143333330342	54
G	33134223225353323444	61
T	44441214430111134110	44

Figure 4.1a Gel mobility shift experiments used to purify random oligonucleotides containing CREA binding sites (CASTing)

Autoradiograph of the gel mobility shift assay showing:

First round of selection. The pool of labelled oligonucleotides 687 (2 μg) was incubated with 75 μg GST-CREA fusion protein (**Tracks 1 & 2**) and with 200 mM EDTA (**Track 3**).

Second round of selection. Oligonucleotides excised from **Box A** were eluted, PCR amplified, and incubated with 75 μg GST-CREA fusion protein (**Tracks 5 & 6**) and with 200 mM EDTA (**Track 4**).

Third and final round of selection.

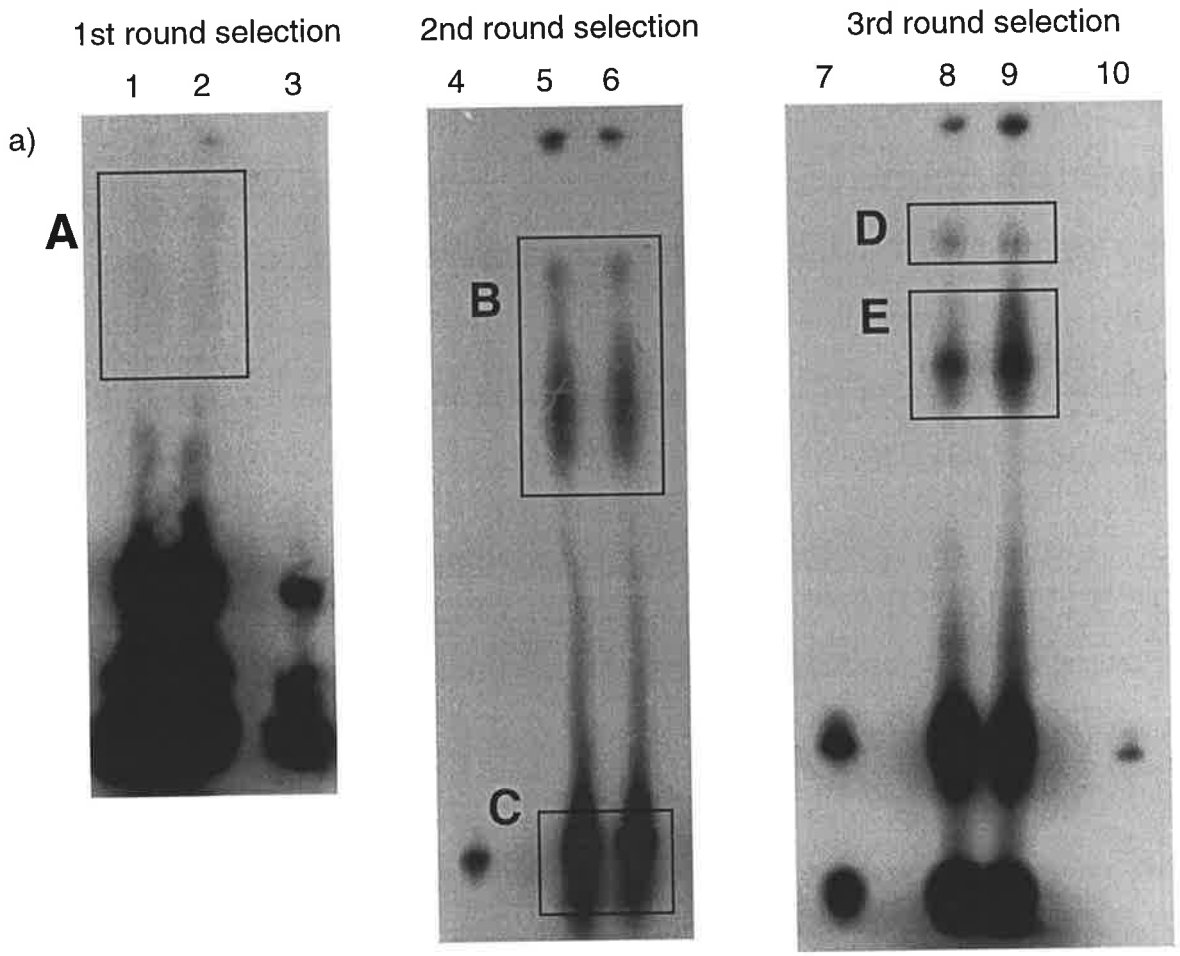
Oligonucleotides from **Box B** were excised, PCR amplified, incubated with 75 μg GST-CREA fusion protein (**Track 8**) and with 200 mM EDTA (**Track 7**).

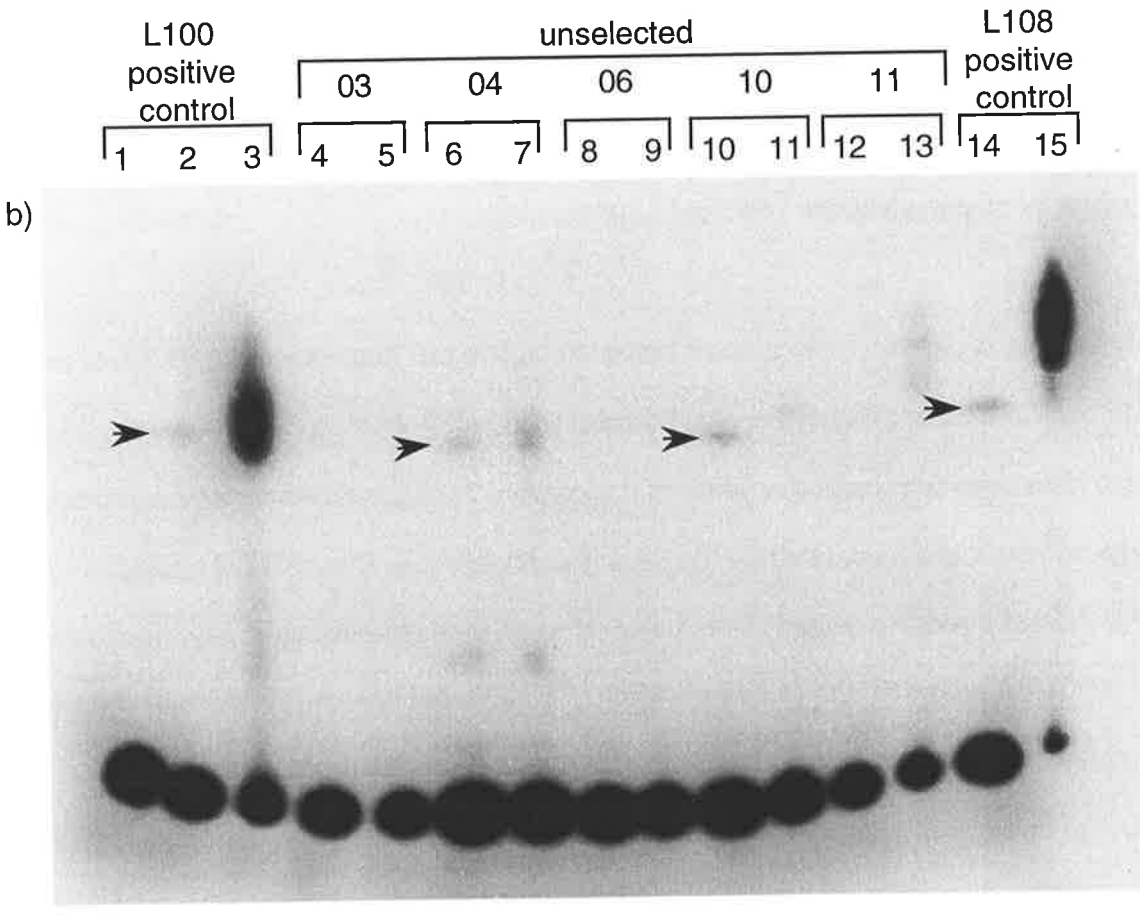
Oligonucleotides from **Box C** were excised, PCR amplified, incubated with 75 μg GST-CREA fusion protein (**Track 9**) and with 200 mM EDTA (**Track 10**).

Oligonucleotides from **Box D** were called the "L" oligonucleotides and oligonucleotides from **Box E** were called the "S" oligonucleotides.

Figure 4.1b Gel mobility shift assay of unselected oligonucleotides of the 687 pool

Gel mobility shift assay of labelled oligonucleotides, unselected (Table 4.1) and "selected" (Appendix A) oligonucleotides incubated with PBS (**Track 1**), 150 μg total protein extract containing GST (**Tracks 2, 4, 6, 8, 10, 12 & 14**) or 80 μg total protein extract containing GST-CREA (**Tracks 3, 5, 7, 9, 11, 13 & 15**). The arrows indicate the position of the complex formed with the GST containing extract.


1st round selection
2nd round selection
3rd round selection
1 2 3
4 5 6
7 8 9 10
a)
A
B
C
D
E


L100 positive control
unselected
L108 positive control
03
04
06
10
11
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
b)

of binding CREA. Both of these DNA samples were amplified and used again in a mobility shift assay. The low mobility complex was clearly made up of two separate complexes, "L" (Box D) and "S" (Box E), which were excised separately and reamplified. In order to purify 50 bp DNA fragments the oligonucleotides were size fractionated on a 4% Metephor agarose gel, electrophoresed onto DEAE membranes and eluted. The oligonucleotides were cloned directly into the vector pGEM-T (Promega), and 48 clones from the low mobility ("L") and 44 clones from the higher mobility ("S") complexes were chosen at random and sequenced. The raw sequence data is presented in Appendix A.

4.1.3 Gel mobility shift assays of unselected and "selected" oligonucleotides

To determine to what degree the GST-CREA fusion protein was able to bind oligonucleotides of this length non-specifically plasmid clones containing unselected oligonucleotides were subject to PCR (section 2.2.8.3). The oligonucleotides were purified by electrophoresis on a 10% non denaturing polyacrylamide gel, excised and eluted. They were tested for their ability to bind CREA in a gel mobility shift assay. The results (Fig. 4.1b) show that these unselected oligonucleotides do not bind CREA. Some do bind a protein present in some *E. coli* extracts, indicated by the arrows. Upon long exposure, there is evidence that oligonucleotide 11 does produce a small amount of binding to GST-CREA which is much less significant than the binding by oligonucleotides L100 and L108 (see below).

All sequenced oligonucleotides were tested to confirm that they bound GST-CREA, and that they were not an artefact of the experimental protocol. For this, plasmid clones of selected oligonucleotides were amplified using PCR (section 2.2.8.3), purified on 10% non-denaturing polyacrylamide gels, and tested for their ability to bind CREA. Clone 03 (as a negative control) and L108 (as a positive control) were included on all mobility shift gels. All oligonucleotides bound CREA except for the 11 listed in Table 4.2. A continuum of affinities was observed for the other 75 oligonucleotides. These were assigned scores of -/+ (weak), + (average) or ++ (strong) binding in Appendix A, relative to 03 and L108 for their affinity to the GST-CREA

Table 4.2 Oligonucleotides that upon testing do not bind CREA appreciably

Name	Sequence
L102	AACGAATTGAACGCGCGGTCGTGACATCATGGCGGGTTACACCATTGA
L104	AACGAATTGAACGCGCCCCTTCTCATAGACTGCGGGTTACACCATTGA
L109	AACGAATTGAACGCCCAACGCTGATCCATGCCGGGGTTACACCATTGA
L121	AACGAATTGAACGCCACATTTGTCCCTCGCCGCGGGTTACACCATTGA
L122	AACGAATTGAACGCGGCCGAGATGGTCATAGTGGGGTTACACCATTGA
L138	AACGAATTGAACGCGGTGCCATGGGACATTTGCAGGTTACACCATTGA
L149	AACGAATTGAACGCCGACATGGAGGGAGCTGGGGGGTTACACCATTGA
S101	AACGAATTGAACGCTACTATAGACGCGGTATAGGGGTTACACCATTGA
S104	AACGAATTGAACGCGGAGTACGCCTACAGCAGCGGGTTACACCATTGA
S136	AACGAATTGAACGCCAACGAGTCACTCCCTGCTAGGTTACACCATTGA
S138	AACGAATTGAACGCCACGTTTTCTGGGTGCCCGGGGTTACACCATTGA

Table 4.3 Oligonucleotides that upon testing bind CREA weakly

Name	Sequence	Mobility
L105	AACGAATTGAACGCCACGAGCGGGTATAGCCTGGGGTTACACCATTGA	HMC
L111	AACGAATTGAACGCCGGATGGGGAAGTAGCGTGGGTTACACCATTGA	HMC
L115	AACGAATTGAACGCCCGACTTTTGTTCATCGTGGGGTTACACCATTGA	HMC
L116	AACGAATTGAACGCGGCCCTCATGTAGATCCCGGGGTTACACCATTGA	HMC
S117	AACGAATTGAACGCGGGTGAAGTATCATTCGCGGGTTACACCATTGA	LMC

fusion protein (Fig. 4.2a). This score was determined by estimating the percentage of the oligonucleotide that was bound to that which was free. In addition, the oligonucleotide-GST-CREA complexes varied in electrophoretic mobility. Higher mobility complexes (HMC) and lower mobility complexes (LMC) were distinguishable and repeatable. A mobility shift assay illustrating this difference is shown in Figure 4.2b. Mobilities are indicated in Appendix A as either HMC or LMC.

4.2 Sequence analysis of the "selected" oligonucleotides

Of the 92 clones sequenced, 85 contained single inserts of 47-48 bp and seven either had no insert or contained deletions of at least one of the sequencing primer sites. Of the 86 clones containing inserts 11 do not bind CREA appreciably, some of these do not contain a string of 4G's (Fig. 4.2c) whereas others did (Table 4.2). All clones that bound with average (+) and strong (++) affinity contained a string of 4G's. In further analyses, this sequence was used as an anchor in alignments to determine whether other positions relative to these 4G's showed a consensus. Of the five clones that bound with weak (-/+) affinity (Table 4.3) there was one, S117, that does not contain a 4G string. These five clones were not included in the alignment analyses discussed below but are considered further in sections 4.2.3-5.

4.2.1 Are there differences between "S" and "L" groups?

"S" and "L" groups contain oligonucleotides with the whole range of binding affinities and there are only minor sequence biases (Appendix B, see also Table 4.5). Whether an oligonucleotide formed a LMC or a HMC did not correlate with the "S" and "L" groupings and there was no difference in the number of sets of 4G's present. This cannot be due simply to errors when cutting out the two complexes. The mobility of LMC and HMC oligonucleotides corresponds to a region within the area cut out as "S". This means that the "L" complex is an artefact of the experiment but that within the "S" region there are oligonucleotides that migrate slower (LMC) and those that migrate faster (HMC). The presence of the "L" complex may reflect a low level of dimerisation between fusion protein molecules. This may be the explanation for the third

band (labelled DC= dimerisation complex) present in the mobility shift assay shown in Figure 3.3c and may be due to the GST portion of the fusion protein. The "S" and "L" groupings thus do not represent different mobility clones and the results of both "S" and "L" clones have been pooled for further analyses.

4.2.2 Is there an extended consensus?

In order to identify any extended consensus sequence that may be present for oligonucleotides that form a HMC with CREA, an alignment was made of the 49 oligonucleotides that bind with average (+) or strong (++) affinity (Table 4.4). Oligonucleotides with more than one string of 4G's were omitted since it was not possible to know which string to align. The number of oligonucleotides with a string of 5G's or more was approximately one quarter of the total number. Therefore a string of 5G's was considered as a string of 4 with the most 5' G forming the most 5' G of the 4G string (position 4, see below). The total numbers for each base at each position relative to G1 the most 3' G in the 4G string are shown in Figure 4.3. A requirement for a C or T at position 5 and a G or C at position 6 is in keeping with previous findings of an SYG GGG consensus for CREA (discussed in sections 3.8-3.9). A T at position 5 was twice as frequent as a C. A G at position 6 was ten times as frequent as a C. Further 5' to this, two and a half times as many oligonucleotides contain a T in position 7 compared to all other bases. An A in positions 5-7 or a T in position 6 was rarely found. Positions 8-19 are AT rich (58%) but there was no obvious requirement for an A or a T at any one particular position. Position 17 has an AT frequency twice that of the number of G's and C's and position 18 is the first position where the frequency of a T falls well below average. In position 20 there is a preference for a C and not an A and the absence of a T in position 21 is very marked. Although five of the C's in position 22 belong to the primer sequence this position is extremely biased toward a C. In all cases position 23 contains a base from the primer 602.

Since the majority of oligonucleotides contain the 4G string adjacent to primer 603 it is not possible to identify 3' preferences.

Figure 4.2a Gel mobility shift assay of "selected" oligonucleotides which display varying degrees of binding affinity to the GST-CREA fusion protein

Gel mobility shift assay of labelled oligonucleotides, selected as able to bind GST-CREA (Appendix A) incubated with PBS (**Track 1**), 150 μg total protein extract containing GST (**Tracks 2, 4, 6, 8, 10, 12 & 14**) or 80 μg total protein extract containing GST-CREA (**Tracks 3, 5, 7, 9, 11, 13 & 15**). The name and assigned score for affinity for each oligonucleotide is also shown.

Figure 4.2b Gel mobility shift assay using the GST-CREA fusion protein and "selected" oligonucleotides which display varying mobilities upon electrophoresis

Gel mobility shift assay of labelled oligonucleotides, selected as able to bind GST-CREA (Appendix A) incubated with PBS (**Track 1**), 150 μg total protein extract containing GST (**Tracks 2, 4, 6, 8, 10 & 12**) or 80 μg total protein extract containing GST-CREA (**Tracks 3, 5, 7, 9, 11 & 13**). The name and assigned score of affinity for each oligonucleotide is also shown.

Figure 4.2c Gel mobility shift assay of "selected" oligonucleotides lacking a GGGG sequence

Gel mobility shift assay of labelled oligonucleotides, selected as able to bind GST-CREA (Appendix A) but lacking a 4G string, incubated with PBS (**Track 1**), 150 μg total protein extract containing GST (**Tracks 2, 4, 6, 8, 10, 12 & 14**) or 80 μg total protein extract containing GST-CREA (**Tracks 3, 5, 7, 9, 11, 13 & 15**). The name and assigned score of affinity for each oligonucleotide is also shown.

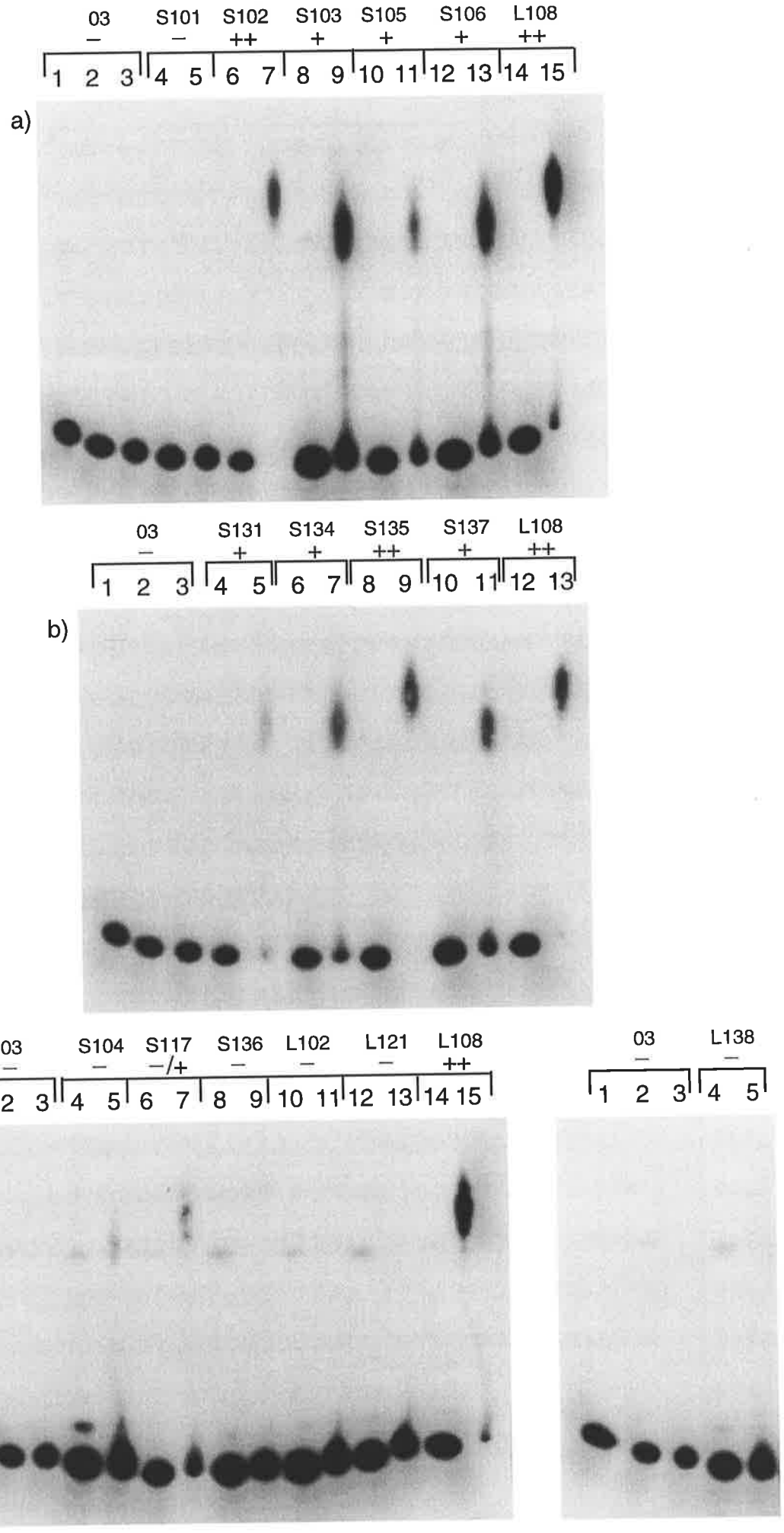

03 S101 S102 S103 S105 S106 L108
– – ++ + + + ++
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
a)
03 S131 S134 S135 S137 L108
– + + ++ + ++
1 2 3 4 5 6 7 8 9 10 11 12 13
b)
03 S104 S117 S136 L102 L121 L108
– – –/+ – – – ++
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
c)
03 L138
– –
1 2 3 4 5

Table 4.4 Alignment of oligonucleotides* anchored by the string of 4G's

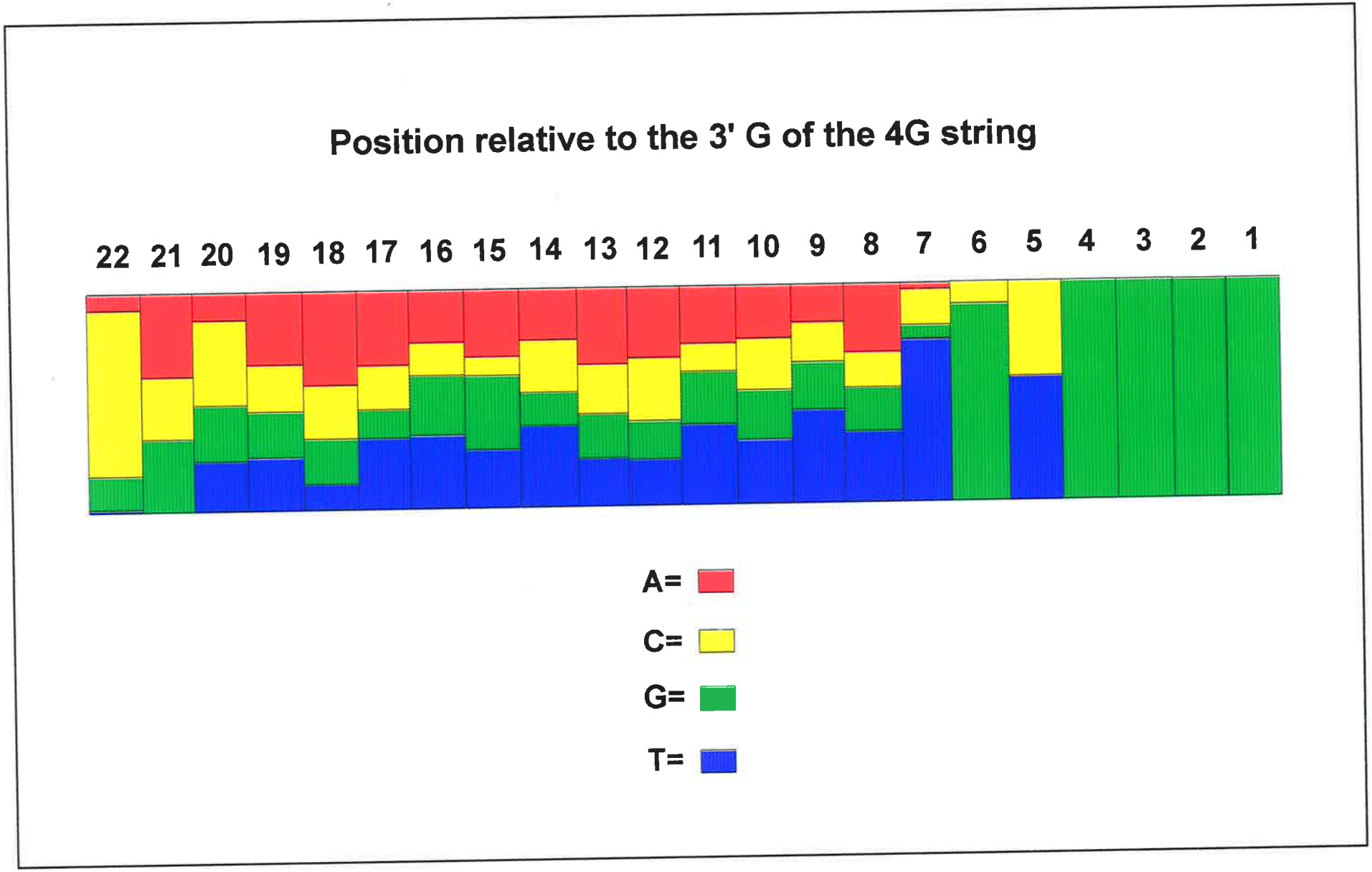
Name	Sequence
L100	AACGAATTGAACGC**CATATACAACGCGGGTGTGG**GGTTACACCATTGA
L101	AACGAATTGAACGC**CGCGACGATCATATTTGTGG**GGTTACACCATTGA
L103	AACGAATTGAACGC**CCATGTTATATATCCTGTGG**GGTTACACCATTGA
L106	AACGAATTGAACGC**CATCCGCATTTAGTCTGTGG**GGTTACACCATTGA
L107	..AACGAATTGAACGC**CACCGGGTCAAATTGTGGGG**GGTTACACCATTGA
L110	AACGAATTGAACGC**CATTGTAGAGCACATCCCGG**GGTTACACCATTGA
L112	AACGAATTGAACGC**CCCAGTTCGACCTTCCCCGG**GGTTACACCATTGA
L114	AACGAATTGAACGC**CACGTATATGGTCCGCGCGG**GGTTACACCATTGA
L115	AACGAATTGAACGC**CCGACTTTTGTTCATCGTGG**GGTTACACCATTGA
L116	AACGAATTGAACGC**GGCCCTCATGTAGATCCCGG**GGTTACACCATTGA
L118	AACGAATTGAACGC**CCGAGGTATACTTAATGCGG**GGTTACACCATTGA
L119	AACGAATTGAACGC**GGCAGGGGTTGCTTATGCGG**GGTTACACCATTGA
L123	AACGAATTGAACGC**CACAACTTGCTATTTTGTGG**GGTTACACCATTGA
L125	AACGAATTGAACGC**CATAACAGAAGATTACGCGG**GGTTACACCATTGA
L126	.AACGAATTGAACGC**CACGTTACCATATCGGCGG**GGTTACACCATTGA
L127	AACGAATTGAACGC**CACTGTGGATCCGCGTGTGG**GGTTACACCATTGA
L128	AACGAATTGAACGC**CATGCTCGCAAACTAAGTGG**GGTTACACCATTGA
L129	AACGAATTGAACGC**CACAAAGAGCAATCATGCGG**GGTTACACCATTGA
L131	AACGAATTGAACGC**GCCTTCTCCTTGTTCCGTGG**GGTTACACCATTAA
L132	.....AACGAATTGAACGC**CATGTGGGGGTCCTTGGCA**GG
L134	..TCAATGGTGTAACC**CGCACGTGAAAAAATGTGGG**GCGTTCAATTCGTT
L140	.AACGAATTGAACGC**CGCCATATGTTTTTTCCGG**GGTTACACCATTGA
L141	AACGAATTGAACGC**CACGCAACCGCTAGATGTGG**GGTTACACCATTGA
L142	AACGAATTGAACGC**CATGACAATTCTGGTTGCGG**GGTTACACCATTGA
L148	ACGAATTGAACGC**CGTGCGGACCGTGGCTGTGG**GGTTACACCATTGA
L150	AACGAATTGAACGC**GCCCAACTTAATATTCCTGG**GGTTACACCATTGA

S103	AACGAATTGAACGC**CACGTTACTATAGCTTGCGG**GGTTACACCATTGA
S105	AACGAATTGAACGC**CGTCGAGGTGAGGGATGTGG**GGTTACACCATTGA
S106	AACGAATTGAACGC**CAGGAGAACGCATTCTGTGG**GGTTACACCATTGA
S107	AACGAATTGAACGC**CGCGAGGTCTGCCCATGTGG**GGTTACACCATTGA
S108	AACGAATTGAACGC**CGGCATTGGACTGGGTGTGG**GGTTACACCATTGA
S109	AACGAATTGAACGC**CACTGATCGTAAGTGTGTGG**GGTTACACCATTGA
S110	..AACGAATTGAACGC**CATCTCTAGTGAGTTGCGGG**GGTTACACCATTGA
S111	AACGAATTGAACGC**GCAAAAGTCCAGCTGTGTGG**GGTTACACCATTGA
S112	AACGAATTGAACGC**CACAATTGCAGTTTGGGTGG**GGTTACACCATTGA
S113	AACGAATTGAA.CGC**GGGTAAGATTGTGATCTGG**GGTTACACCATTGA
S118	AACGAATTGAACGC**ACCTGATGCAACTTTCGCGG**GGTTACACCATTGA
S120	.AACGAATTGAACGC**GGACTATGCTGAGATGTGG**GGGTTACACCATTGA
S123	.............AACGAATTGAACGCC**ACGCGGGGTATTAGGACAC**GG
S124	AACGAATTGAACGC**GGGAGATGACGTCTGTGTGG**GGTTACACCATTGA
S129	AACGAATTGAACGC**GGCCCTAGTACCCTATGTGG**GGTTACACCATTGA
S131	AACGAATTGAACGC**CCGTACGTTGCCCATTGCGG**GGTTACACCATTGA
S132	AACGAATTGAACGC**GGAAAGGTTACTCACTGTGG**GGTTACACCATTGA
S134	AACGAATTGAACGC**CGTTAACTTCATGAATGTGG**GGTTACACCATTGA
S137	AACGAATTGAACGC**CATCCCGGTAATATTCGTGG**GGTTACACCATTGA
S139	AACGAATTGAACGC**CGCACATACTAGATTGGTGG**GGTTACACCATTGA
S141	AACGAATTGAACGC**CACCGAATAGATTTGTGTGG**GGTTACACCATTGA
S142	AACGAATTGAACGC**TGATAAAAATAGTGGTGCGG**GGTTACACCATTGA
S143	AACGAATTGAACGC**CGGTATCTTCGGAATTGTGG**GGTTACACCATTGA

*all oligonucleotides forming a HMC with GST-CREA and containing a string of 4G's

Figure 4.3 Diagram of the numbers of each base present at each position for oligonucleotides binding with + or ++ affinity forming a HMC

Oligonucleotides from Table 4.4 were aligned by the 4G string and the number of bases (A, C, G and T) at each position 5' of this sequence is shown. In all but five cases the G1 and G2 bases are specified by the primer 603 sequence. Raw data appear in Appendix B (Table B.3).


Position relative to the 3' G of the 4G string
22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
A=
C=
G=
T=

4.2.3 Sequence comparison of oligonucleotides that bound well with those that bound poorly or not at all

In order to determine what distinguishes oligonucleotides that bind from those that bind poorly (Table 4.3) or not at all (Table 4.2), oligonucleotides from Table 4.2, 4.3 and 4.4 were analysed according to the 3 bases 5' of the 4G string (only those oligonucleotides containing a 4G string were included). This is shown in Table 4.5.

In agreement with the individual preferences the most common triplet was TGT. There were no oligonucleotides containing the three most common triplets (TGT, TGC and CGC) that did not retard. This reflects the importance of G6 in binding affinity. There were however two oligonucleotides with a G6 that did not retard. One of these L122 contained an AGT triplet, as did one oligonucleotide which bound (L128). Their sequences are:

L128 AACGAATTGAACGCCATGCTCGCAAACTA**AGT**GGGGTTACACCATTGA

L122 AACGAATTGAACGCGGCCGAGATGGTCAT**AGT**GGGGTTACACCATTGA

Within the region (underlined) where the two oligonucleotides differ, L128 has a C in position 22 which was consistently preferred in oligonucleotides that bound well. Furthermore, L128 contains 4 A's and a T within positions 8-13, and specifically contains a T in position 9 which almost half the oligonucleotides from Table 4.4 contained. L122 does not have a C22 nor a very AT rich region 5' of the AGT GGGG sequence. The other oligonucleotide that contains a G6 and did not bind well was L115. It contains a CGT triplet which is also present in two oligonucleotides that showed binding, namely S137 and L131. Their sequences are:

L131 AACGAATTGAACGCGCCTTCTCCTTGTTC**CGT**GGGGTTACACCATTAA

S137 AACGAATTGAACGCCATCCCGGTAATATT**CGT**GGGGTTACACCATTGA

L115 AACGAATTGAACGCCCGACTTTTGTTCAT**CGT**GGGGTTACACCATTGA

The two oligonucleotides that bind well both contain a T9 whereas L115 does not. S137 contains a C22 and six A's and T's in positions 8-13. L115 contains a C22, but only 4 A's and T's in positions 8-13, which is similar to L131. T9 may be important for oligonucleotides with poorer triplets 5' of the 4G string.

Table 4.5 Sequence of 3 bp 5' of the 4G string present in oligonucleotides from Tables 4.2, 4.3 and 4.4

Triplet 5' of the 4G string	Number of oligonucleotides				
	bind well			weakly	not at all
	total	"L"	"S"		
TGT	24	11	13		
TGC	7	3	4		
CGC	4	2	2		
CGT	2	1	1	1	
GGT	2		2		
GGC	1	1			
AGT	1	1			1
TCT	1		1		
TCC	1	1			
CCT	1	1		1	
CCC	1	1		1	1
GAT				1	
GCT					1
GCC					1
ATA					1
AAG					1

There were only 4 oligonucleotides that bound well that did not contain a G6. Two of these, S113 and L140 have a T7 (TCT and TCC respectively). There were no oligonucleotides that showed weak or absence of binding that contained these triplets thus a comparison is not possible. The third oligonucleotide that does not contain a G6 as part of the triplet (CCT) is L150. This triplet occurs within L105, a weakly binding oligonucleotide and their sequences are:

L150 AACGAATTGAACGCGCCCAACTTAATATT**CCT**GGGGTTACACCATTGA

L105 AACGAATTGAACGCCACGAGCGGGTATAG**CCT**GGGGTTACACCATTGA

L150 contains a very AT rich region 5' of the core recognition site including aT9 whereas L105 does not. Perhaps the presence of a C22 has allowed weak binding by CREA for the triplet contained in L105 but strong binding was perhaps only possible with an AT rich region. The AT rich region present in the MIG1 consensus was shown to bend and thus bending may be of importance here. The forth oligonucleotide that binds but does not have a G6 is L110. It has a CCC triplet and so do the weakly binding L116 and the non binding S138. Their sequences are:

L110 AACGAATTGAACGCCATTGTAGAGCACAT**CCC**GGGGTTACACCATTGA

L116 AACGAATTGAACGCGGCCCTCATGTAGAT**CCC**GGGGTTACACCATTGA

S138 AACGAATTGAACGCCACGTTTTCTGGGTG**CCC**GGGGTTACACCATTGA

The non binding oligonucleotide does not contain a very AT rich region 5' of the core. This may be a significant factor for this triplet. The L116 oligonucleotide does not have a C22 but does have a very GC rich region (eight straight G or C bases). It is possible that higher affinity binding by L110 is due to the presence of a C22 and the AT rich region.

It must be mentioned that the only HMC oligonucleotide with strong affinity was L134 and this oligonucleotide contains a string of six A's 5' of the core T GTG GGG recognition sequence. The results presented above indicate that the oligonucleotides selected in this way displayed quite strong sequence preferences for positions 3-8 and also for positions 21-22. Due to the short length of the random region (20 bp) it is not possible to say how much of the 602 primer sequence contributes to recognition. Such marked biases at both ends of the random sequence

implies that GST-CREA recognises two distinct regions of the DNA for good binding. These regions were widely spaced at two turns of the helix. The primer 602 sequence of CGC may not be optimal but may allow good binding due to its GC richness and because eight out of the eleven bases 5' of this are A or T. Many oligonucleotides in Table 4.4 contain GC rich regions at the opposite end to the binding site often up to 6 bp in length. An A in position 21 was often found but a T was not.

4.2.4 What are the sequence differences between LMC and HMC forming oligonucleotides?

The ability of some oligonucleotides to form HMC's and others LMC's with GST-CREA was clearly distinguishable on mobility shift gels. Labelled L108, generated many times from different PCR reactions, never varied with respect to its lower mobility as a complex with CREA. However since oligonucleotides were generated by PCR amplification it was theoretically possible that the HMC and LMC complexes reflected differences in DNA concentration. To test this a gel mobility shift assay was carried out where L108 was incubated with varying amounts of GST-CREA fusion protein (Fig. 4.4a). A two fold increase or decrease in protein concentration did not change the mobility of the DNA-CREA complex. A small amount of GST-CREA protein did not produce both a HMC and an LMC but rather a smaller percentage of the complex was present as a LMC and the rest was present as unretarded probe. Therefore it seems likely that whether a LMC or a HMC is formed depends on the sequence of the DNA and does not reflect large differences in DNA concentration.

Twenty three oligonucleotides formed a LMC (Table 4.6). A likely explanation for LMC formation is the binding of additional CREA protein molecules to the HMC oligonucleotide. Nine LMC forming oligonucleotides contain two sets of 4G's. Binding to S117 by CREA, albeit with low affinity, results in a LMC and might therefore, contain two binding sites. Present in the S117 sequence is a direct repeat GCG GGT which differs from the SYG GGG consensus for CREA at only one position resulting in a sequence not previously identified as a recognition

sequence. Therefore it is possible that the LMC group is made up of inserts with more than one binding site for CREA where one contains 4G's and the other is new or some variation on the 4G sequence. Analysis of the other LMC sequences revealed that seven contained a string of 4G's together with a GYG GGT sequence, three contain a string of 4G's together with a SYG GAG sequence and one (S133) contains a GCG GGC sequence and a string of 4G's. Two exceptions, S119 and S121, contain a set of 4G's but no additional string of 3 or 4 G's. S119 has a CGCC and a CAT GGT sequence and S121 contains a CGCC sequence or a GTG GCG sequence on the other strand. S119 bound with strong affinity while S121 bound with only average affinity.

The three bases 5' of the 4G strings of the LMC oligonucleotides (Table 4.6) include the triplets most commonly found in oligonucleotides that form a HMC with average to strong binding namely TGT (7), TGC (3) and CGC (10). All LMC forming oligonucleotides contain triplets with at least one G, most commonly this was a G6. The number of CGC triplets is probably over-inflated since all of these are present in the primer 602 sequence. All oligonucleotides containing at least one of these common triplets in front of a 4G string bound with strong affinity. Many oligonucleotides contain one triplet as above plus one containing A's in positions 5-7 or lacking G6's or T7's. The presence of these other triplets not found commonly in oligonucleotides from Table 4.4 indicates that co-operative binding by CREA may be occurring. In this way a neighbouring good site decreases the activation energy required to occupy sites of poorer affinity. Only one third of the CREA protein was expressed as a fusion protein so it is likely that the CREA-CREA protein interaction domains are not present. However, the GST portions of the fusion proteins are able to interact with each other and this may stabilise the binding of CREA to a poorer site either by increasing the local concentration of CREA or by bringing into close proximity some region of the CREA protein which forms an interaction sufficient to result in co-operative binding.

4.2.5 What sequence preferences are there for a GYG GGT string?

By comparing the GYG GGT sequence found within oligonucleotides that form a LMC to those within oligonucleotides that form a HMC, it was possible to determine which bases were

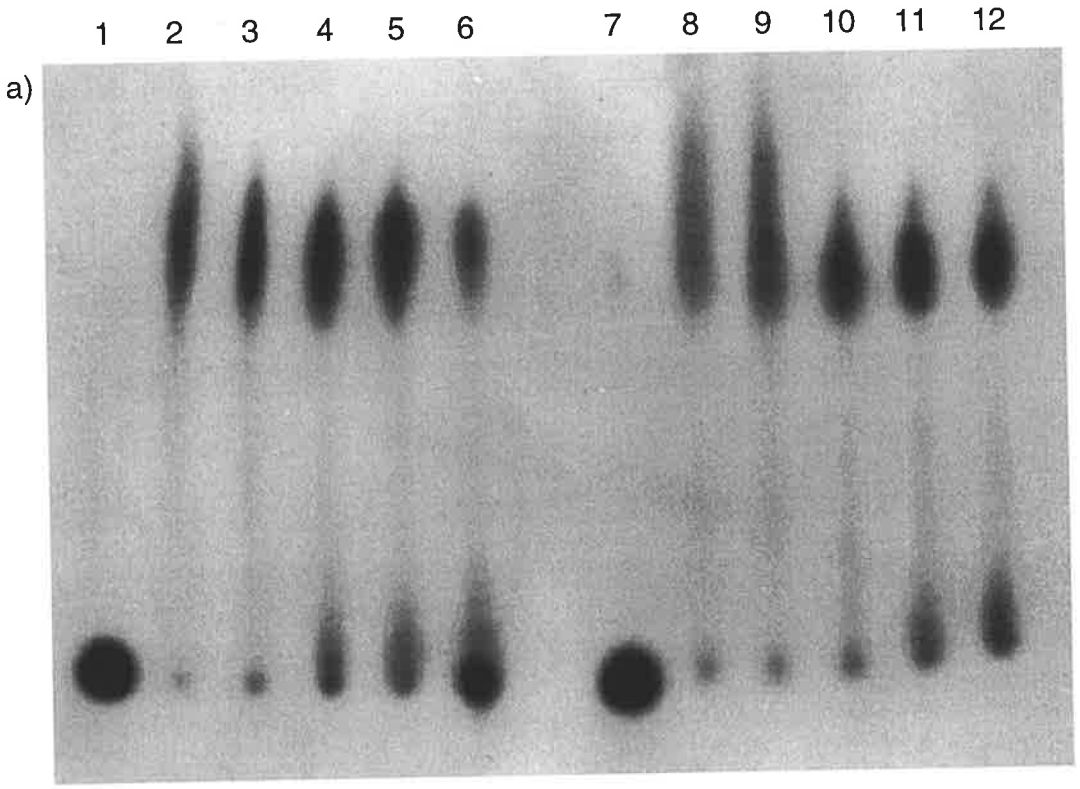
Figure 4.4a Gel mobility shift assay of a LMC and HMC forming oligonucleotide with varying amounts of GST-CREA fusion protein

Gel mobility shift assay of labelled oligonucleotides L108 (**Tracks 1-6**) and L134 (**Tracks 7-12**) incubated with PBS (**Track 1 & 7**) or 300 μg total protein extract containing GST-CREA (**Track 8**) or 150 μg total protein extract containing GST-CREA (**Track 2 & 9**) or 75 μg total protein extract containing GST-CREA (**Track 3 & 10**) or 30 μg total protein extract containing GST-CREA (**Track 4 & 11**) or 15 μg total protein extract containing GST-CREA (**Track 5 & 12**) or 1.5 μg total protein extract containing GST-CREA (**Track 6**).

Figure 4.4b Results of methylation and depurination interference assays using the GST-CREA fusion protein and the 3.1-3.3 ***prn*** **intergenic region**

Methylation (A) and depurination (B) interference footprints of the region of DNA spanning the **3.1**, **3.2** and **3.3** CREA binding sites (shaded boxes) within the *prn* intergenic region. Left panel, coding strand; right panel, non-coding strand. Black dots represent 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. fp, free probe, no protein added; b, bound probe, 100 ng of fusion protein added. Shown is Figure 8 from Cubero and Scazzocchio, (1994).

a)


1 2 3 4 5 6 7 8 9 10 11 12

b)

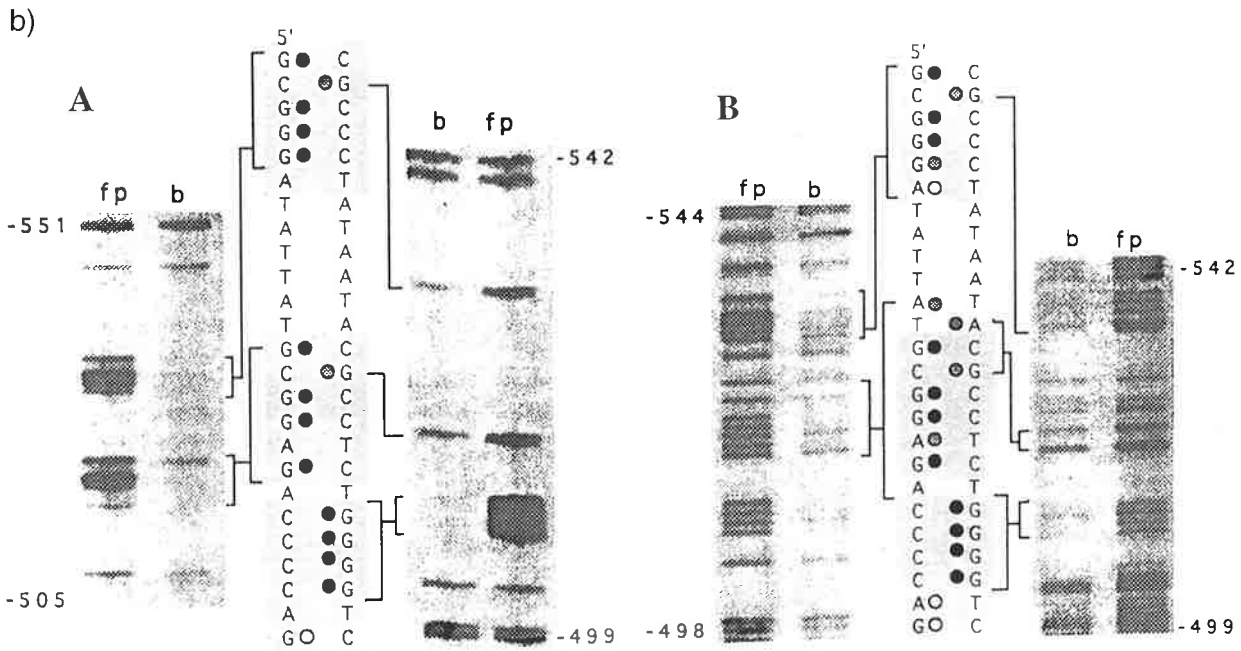

A
fp
b
-551
-505
5'
b
fp
-542
-499
B
fp
b
-544
-498
b
fp
-542
-499

Table 4.6 Oligonucleotides that form a LMC when bound to GST-CREA

Name	Sequence	Affinity
L108	AACGAATTGAACGCGGGGTGCGCTACGTATGGCGGGTTACACCATTGA	++
L124	AACGAATTGAACGCGGAGGAGATTGTGGGGGATGGGTTACACCATTGA	++
L133	AACGAATTGAACGCTGGCGGGGGTCTCACAGTGGGGTTACACCATTGA	++
L137	AACGAATTGAACGCCCCAGACCCCGTACTTGCCAGGTTACACCATTGA	+
L139	AACGAATTGAACGCGGGTGGTCGTGACACGTTGGGGTTACACCATTGA	+
L144	AACGAATTGAACGCGGGGACAGTTGCTTTTGTGGGGTTACACCATTGA	++
L145	AACGAATTGAACGCGGGGGTATGCGGTGTGCTGGGGTTACACCATTGA	++
L147	AACGAATTGAACGCCATGGAGGGGGACTTGCGGGGGTTACACCATTGA	+
S100	AACGAATTGAACGCGGAGAGCATATATGGCCGGGGGTTACACCATTGA	+
S102	AACGAATTGAACGCGGGGTGCTCTCGTTGTGCGGGGTTACACCATTGA	++
S115	AACGAATTGAACGCGGGGGCACACTTTTGCAAGGGGTTACACCATTGA	++
S117	AACGAATTGAACGCGGGTGAAGTATCATTCGCGGGTTACACCATTGA	-/+
S119	AACGAATTGAACGCGGGGGTCGCCTGAGTTGCATGGTTACACCATTGA	++
S121	AACGAATTGAACGCCACTTGATGTGAGACTGTGGGGTTACACCATTGA	+
S122	AACGAATTGAACGCGGAGGGCGCTGTACGTGTGGGGTTACACCATTGA	++
S125	AACGAATTGAACGCACGTTGGGGCGAGTATGTGGGGTTACACCATTGA	++
S126	AACGAATTGAACGCGGGGAAATGGTATAATGGCGGGTTACACCATTGA	++
S127	AACGAATTGAACGCGGTATGTGGGTTTTATGTGGGGTTACACCATTGA	++
S128	AACGAATTGAACGCGGGGGGGGGCGCGTCTCGTGGGTTACACCATTGA	+
S130	AACGAATTGAACGCGGGGCAGTTACTCATGGCGGGTTACACCATTGA	++
S133	AACGAATTGAACGCGGGCGACTACGCTTATGTGGGGTTACACCATTGA	++
S135	AACGAATTGAACGCGGAGTACGAGGGGATTGCGGGGTTACACCATTGA	++
S140	AACGAATTGAACGCGGGGGGGATTTTGTTTGGCGGGTTACACCATTGA	++

important for binding. Both GTG GGT and GCG GGT sequences were present in LMC oligonucleotides. In the absence of a G6 (numbered from the 3' T), the oligonucleotide migrated as a HMC (eg L105 and S113). Although the number of GYG GGT containing oligonucleotides is small, an alignment of their 5' sequences indicates that an extended consensus of A/T A/T N G C/T G GGT exists (Table 4.7). An oligonucleotide (L111) that does not fit this extended consensus and has a GAT triplet 5' of the 4G string, migrates as a HMC.

The importance of the G6 is evident since it is required in the GYG GGT string and is preferred in the 4G string also. The only two oligonucleotides that do not have at least one 4G string with a G6 (ie L137 and L139) both form LMC's with only average affinity.

4.2.6 The importance of orientation

There was no example of a LMC with a string of 4G's and a string of 4C's on the same strand (ie the G's from both binding sites were always on the same strand of DNA). Some oligonucleotides (L104, L112 and L134) contain two binding sites on opposite strands of the DNA but still migrate as a HMC. This may reflect the importance of orientation to binding.

4.3 Discussion

An extended 5' consensus for oligonucleotides that bound CREA with average to strong affinity and containing only one string of 4G's was determined. Within a core SYG GGG sequence, a preference for a T in position 5 (two and a half times) and a G in positon 6 (ten times) was favoured over a C in either position. There was a preference for an AT containing region 5' of this until position 20-22 where oligonucleotides became GC rich. A CGCC sequence was present in all but 11 oligonucleotides that bound with average to strong retadation. There was an absence of A's in positions 5 and 6 and an A in position 7 occured only once. In comparison, of the 11 oligonucleotides that did not retard, 5 did not contain a string of 4G's, three contained A's within positions 5-7 and three contained sequences that must require additional context for binding. These results confirm that MIG1 and CREA recognise very similar sequences.

Table 4.7 5' extended context for the GYG GGT sequence

Name	Sequence
L108	AACGAATTGAACGCGGGGTGCGCTACGTATGGCGGGTTACACCATTGA
L139	AACGAATTGAACGCGGGTGGTCGTGACACGTTGGGGTTACACCATTGA
S117	AACGAATTGAACGCGGGTGAAGTATCATTCGCGGGTTACACCATTGA
S126	AACGAATTGAACGCGGGGAAATGGTATAATGGCGGGTTACACCATTGA
S127	AACGAATTGAACGCGGTATGTGGGTTTTATGTGGGGTTACACCATTGA
S128	AACGAATTGAACGCGGGGGGGGGCGCGTCTCGTGGGTTACACCATTGA
S130	AACGAATTGAACGCGGGGCAGTTACTCATGGCGGGTTACACCATTGA
S140	AACGAATTGAACGCGGGGGGGATTTTGTTTGGCGGGTTACACCATTGA

Base	Position relative to the 3' T													
	14	13	12	11	10	9	8	7	6	5	4	3	2	1
A	1	4	1		2	4	2							
C	2		4	1	1	1		3		6				
G	1	2		4	2			4	8		8	8	8	
T	4	2	3	3	3	3	6	1		2				8
	N	N	N	N	N	A/T	A/T	N	G	C/T	G	G	G	T

4.3.1 A new variation on the recognition sequence for CREA

The finding that S117 bound CREA, albeit weakly, and formed a LMC suggests that the direct repeat GYG GGT is responsible for sequence specific binding to CREA. This sequence was also found in a number of oligonucleotides which formed a LMC but do not contain two sets of 4G's. These results suggest that GYG GGT is a recognition site for CREA when located in close proximity to another site of higher affinity (SYG GGG). This variant shows as similar an affinity to CREA as does the GYG GAG variation found within the *prn* intergenic region (Cubero and Scazzocchio, 1994). Work on MIG1 from *S. cerevisiae* did not identify the GYG GGT sequence as able to bind MIG1. That S119 forms a LMC may be due to binding to a CAT GGT sequence. It is not possible to use DNase I footprinting to determine whether this is the case since the protected regions from the GST-CREA fusion protein were 20 bp long and this is the length of the degenerate region of these oligonucleotides. Chemical footprinting methods for example methylation inference assays would be more informative. This CASTing experiment indicates that CREA cannot bind a totally novel recognition sequence with high affinity, however, the identification of binding sites other than those predicted is significant.

4.3.2 Comparison with other Cys_2His_2 zinc finger proteins

From the specificity of Zif268 and other Cys_2His_2 finger containing proteins (Fig. 3.18a) we can be sure that the -1, and +6 arginine residues from finger I and the -1 arginine from finger II specify a G by hydrogen bonding with the N7 and O6 of the guanine. This interaction is stabilised by the preceding aspartic acid (+2) in all the Zif268 fingers (Pavletich and Pabo, 1991). CREA contains an aspartic acid in this +2 position of finger II but a glutamic acid at +2 of finger I. Aspartic and glutamic acids differ only in a CH_2 moiety and it is therefore likely that glutamic acid (+2) of finger I has a similar function. For Zif268 the histidine specifies a guanine by hydrogen bonding with either the N7 or the O6. CREA contains a histidine at position +3 of ZF I and thus agrees with the results presented here and in chapter 3 which show that G-G GGG is recognised by CREA. Nardelli *et al.*, (1992) show by mutagenesis that a histidine can also specify an adenine. This explains the ability of ZF I of CREA to specify a GAG sequence.

The results presented here suggest that GAG is not as good a target for ZF I of CREA as is the sequence GGG. Cubero and Scazzocchio (1994) report strong binding to this sequence, however, in this experiment no oligonucleotide with this sequence alone showed average (+) or strong (++) binding (see Table 4.4 and S104 in Table 4.2). Therefore the strong binding reported to site **3.2** is probably because this site was located in close proximity to another site with 4G's (**3.1**).

The glutamic acid (+3 fingers I and III) of Zif268 does not make contacts with the recognition sequence however, in this position a C is specified in the Zif268 binding site and also within those of Krox-20 and Sp1 (Christy and Nathans, 1989; Chavrier *et al.*, 1990; Nardelli *et al.*, 1991; Kadonaga *et al.*, 1986, 1987). Comparing the zinc fingers (at the critical positions) of these proteins (Fig. 3.18a) we would accurately predict that the first finger of CREA would recognise G G/A G. This gives some validity to the principle of determining recognition sequence based only on the amino acid residues present at -1, +3 and +6 of the finger. Finger II from CREA would be predicted to specify G/T (C/T/A) G. While position -1 (G) is accurate, the +3 position may be limited in CREA binding to just C/T as an A was not often found. One exception may be binding to the S119 CAT GGT sequence. For the +6 position a T was predicted in addition to the G. This was not observed. The only other base preferred at this position was a C. The model does explain why a G6 is favoured since the Arg to G contact results in two hydrogen bonds being formed. Although a C in position 5 was favoured by MIG1, under the conditions of this experiment a T was preferred (2:1) by CREA. In contrast to what was found with Krox-20 an A was definitely not favourable for binding by CREA or MIG1. Perhaps tight binding by one finger gives the protein some flexibility in the binding of the second finger especially at the end of the recognition site. It is possible that the zinc fingers of CREA do contact the DNA helix in a different way to others in this class and that this produces the deviations observed. There are examples of proteins binding to DNA not through the -1, +3 and +6 positions, for example, the Tramtrack protein from *Drosophila* is thought to make contacts with the +2 amino acids, Ser and Asp (Fairall *et al.*, 1993).

Both fingers of ADR1 have four Arg residues that can bind G, but G is very specifically required by two internal fixed arginines, +6 Arg finger I and -1 Arg finger II (Thukral *et al.*, 1992). In contrast the two external Arg residues -1 (finger I) and +6 (finger II) bind less specifically to G and can be substituted in function by other residues (Taylor *et al.*, 1995). This may explain why a GCG GGT sequence is observed in oligonucleotides forming a LMC (containing more than one site) since it still contains the two fixed internal contacts. An exception to this is the CAT GGT sequence of S119. Furthermore, the +6 Arg of finger II from ADR1 recognises a T and perhaps a similar match is allowed for CREA in position 1 (G1->T1). The flexibility at Arg +6 and to a lesser extent Glu +3 of finger II of CREA for the CREA-DNA interaction may be due to their location at the end of the recognition sequence, especially since a string of 4 G's is present to serve as an anchor.

4.3.3 Comparison with the MIG1 protein from *S. cerevisiae*

MIG1 recognises an SYG GGG core 5' of which is found a T and four bases of either A's or T's. Within this AT rich region Lundin *et al.*, (1994) found that no one particular base was required but the overall AT richness was important, perhaps to allow for bending of the DNA molecule. The results presented here show that a T in position 7 is important for binding by CREA. Some core sequences of the from SYG GGG appeared to require a T9 as well as other sequences up to 12 bp 5'. Although CREA does not require a totally AT rich region 5' of the core sequence a preference for A's and T's over C's and G's was evident. In comparison positions 20-22 were GC rich. Notably the sequence CGCC was present in 60% of oligonucleotides, mostly as an extention of the CGC sequence present in primer 602. Surprisingly the presence of an A or T after the CGC of primer 602 did not abolish binding even though it occured very rarely. Perhaps certain core sequences or an ideal AT rich region obviates the need for a defined CGCC sequence and instead overall GC richness is sufficient. Important to note is that only 11 oligonucleotides that bound CREA did not contain a CGCC sequence or two binding sites. These results do show that GST-CREA is recognising more than just the SYG GGG core sequence previously identified (discussed in sections 3.8-3.9). Recognition of bases up to 12 bp 5' has not been reported for MIG1. In retrospect the choice of primer sequences was

unfortunate in that it is not known how much of the primer sequence contributes to this. The spacing of the GC rich region (including C22) from the SYG GGG core was maximised over the length of the random region. Due to the choice of N=20, it was not determined whether a larger spacing would have been prefered. It is possible that this spacing has been determined not by the CREA portion of the fusion protein but by the GST portion. When two core sequences were present they were most often also at opposite ends of the random region. Perhaps this is because 8 out of 11 bases 5' of the CGC sequence in primer 602 were either A's or T's and therefore this AT richness of the primer sequence may have increased binding affinity. There were oligonucleotides that contained binding sites much closer together and this did not appear to prevent them from forming a LMC.

4.3.4 Presence of core GYG GGG and GC rich sequences in oligonucleotides that do not bind

All the selected oligonucleotides that were subsequently found not to bind CREA significantly have either a 4G sequence or a GYG GGT sequence. In addition all have a very GC rich region next to primer 602 and thus are likely to have a very low level of affinity to CREA. A general attraction to DNA by CREA *in vivo*, especially to GC rich regions, would increase the efficiency of finding target binding sites within chromatin. The fact that all unselected oligonucleotide clones did not bind CREA, except 11, implies that CREA is not able to recognise the primer sequences alone and therefore binding by the selected oligonucleotides is due to the presence of elements within the variable region. Upon long exposure, there is evidence that the unselected oligonucleotide 11 does produce a small amount of binding to GST-CREA. The sequence of oligonucleotide 11 contains a string of 4 G's in it but poor binding might be due to the deviation of the core SYG GGG to an ATG GGG sequence.

4.3.5 Implications for binding to sequences 5' of the genes known to be under CREA control

The work presented shows that a SYG GGG sequence is not sufficient for binding by the CREA fusion protein *in vitro* and is in agreement with the finding that oligonucleotide pair 31/32 did

not bind (section 3.8.2). MIG1, a closely related protein, also requires additional sequence elements for binding, most notably a T7 and an AT rich region. Therefore it is likely that CREA recognises additional elements *in vivo* as well as *in vitro*. In this chapter oligonucleotides with two binding sites were found to have a greater affinity in general than oligonucleotides with only one site and a GC rich region. The region protected in DNase I sensitivity assays of the ALU148 fragment contains many sites which suggests that *in vitro* and *in vivo* this is a strong binding region for CREA. Results presented here also support the model where co-operative binding of CREA to more than one site occurs. Its likely that 31/32 did not retard because a TC CCG GGG binding sequence requires additional elements for strong binding. A similar situation exists for RL2/RL3 which contains one CREA binding site lacking a G6 and another sequence GAG GAG. Therefore the presence of additional binding sites is very important. Similarly, the lack of binding of ONC6/ONC7 to CREA suggests that one divergent CREA site was not sufficient for binding. This work and that on fragments from the 5' regions of *amdS* and *facB* suggest that CREA may require another site in close proximity when a single site diverges from the ideal consensus. This is supported by the recent paper by Espeso and Penalva (1994). These researchers found that divergent sites that bound CREA were always located in close proximity to other sites. These results suggest that additional sequence preferences like CGCC are less important if two binding sites are present. A CGCC sequence was not present in the protected region of ALU148, however, the region around the *Sma* I site is very GC rich. A discrepancy exists since an A5 was not found in oligonucleotides that bound CREA but two of the CREA boxes 5' of *amdS* (**A2** and **A3**) contain an A in position 5. Theoretically we may have expected an A to be acceptible (section 4.3.2) and thus these sites may be functional *in vivo*.

Co-operative binding may explain why the sites in the protected region of ALU148 are divergent from the T GTG GGG preference found here. Co-operative binding would reduce the need for additional protein DNA interactions as the complex is stabilised by protein-protein interactions. Co-operative binding would also have important consequences at the molecular level because it suggests that more than one CREA binding site is present in the 5' region of genes under CREA control, and/or that the more CREA binding sites the tighter the control by

CREA. GYG GGY sites are weak binding sites for CREA as they were only observed within oligonucleotides that contain other sites. A GYG GGY site has not been found within an *in vivo* functional binding site for CREA. However, a T GAG GGT site (on the non-coding strand) lies within the protected region 5' of *amdS* 18 bp downstream of the **A1** site and may be able to be bound by CREA through co-operative binding. A similar sequence, **3.3** (T GCG GGA) was protected in DNase I sensitivity assays and methylation interference studies of the *prn* intergenic region (Cubero and Scazzochio, 1994; see Fig. 3.26). Therefore sites which deviate from the consensus SYG GRG can not be ruled out as being involved in gene regulation *in vivo* especially if they are adjacent to a strong CREA site. Another sequence CTG GGA from the ALU148 fragment is protected on one strand and thus may be involved in the regulation of *amdS* by CREA even though it also deviates from the SYG GRG consensus.

The most popular triplet 5' of the 4G string was determined using CASTing to be TGT. However the majority of sites located 5' of the genes studied did not contain this triplet. Interestingly, if we expand the SYG GRG consensus to include a T or G at position 1 then the palindromic sequence SYG GRS results. CASTing shows that CREA prefers a G in position 6 of the SYG GRG sequence and thus palindromicity was not important.

Cubero and Scazzocchio, (1994), carried out methylation and depurination interference footprinting assays on the DNA fragment which contained the sites **3.1**, **3.2** and **3.3** (Fig. 4.4b; see also Fig. 3.26). These studies confirmed the importance of guanine residues in binding by GST-CREA. Removal of the guanines from the core site on the coding strand (sites **3.2** and **3.3**) or on the non-coding strand (site **3.1**) interfered strongly with binding. In addition removal of guanines which are on the other strand (ie position 6 for site **3.1**; position 5 for sites **3.2** and **3.3**) interfered partially with binding. Interestingly, removal of the adenine at position 2 of site **3.2** resulted in partial interference impling a bond between CREA and this base. One reason why a T in position 7 may be prefered by CREA is that the A on the other strand of site **3.2** is involved in binding. So too was the first A of the AT rich region immediately 5' of the site **3.2**.

These results show that when the first finger is not binding the favourable GGG triplet (as in sites **3.2** and **3.3**) then the G at position 6 and the G on the other strand (equivalent to position 5) is important for binding. The oligonucleotide O3 sequence was based on the sequence of site **3.2** except that the G6 was changed to a C6. This oligonucleotide did not bind GST-CREA at all, consistent with the results of the methylation and depurination interference footprints and with the results obtained here that a G6 is important for binding especially if ZF I is not binding the favourable GGG sequence. When oligonucleotide O2 was tested for binding it was found to bind even better than the original oligonucleotide O1. This oligonucleotide contains a G6, a T5 and a perfect AT rich region 5' of the core GCG GAG binding site. This can be compared to oligonucleotide L134 which contains a GTG GGG sequence, a T7 and a perfect AT rich region 5' of the core sequence. This was the only oligonucleotide which bound GST-CREA with strong affinity yet contained only one site. The reason both of these oligonucleotides bind so well is obviously because they contain many of the right elements within the one oligonucleotide.

4.3.6 More to GST-CREA than just the zinc fingers?

The model that each Cys_2His_2 finger recognises a triplet of bases was first put forward by Nardelli *et al.,* (1991). This model was supported by crystallographic studies of Zif268 (Pavletich and Pabo, 1991). The finding that CREA prefers a T at position 7 and possibly a T9 means that CREA is able to recognise bases further 5' of the predicted 6 bp core sequence, either by interactions with the zinc finger or other portions of the protein (present in the fusion construct). This has not been reported for Zif268, ADR1 or SP1 and thus may mean that CREA (and MIG) interact differently with DNA than does Zif268. Whether CREA requires an AT rich region 5' of the recognition sequence is not really clear. An AT rich region was prefered but not to the extent that it is required for MIG1 binding (Lundin *et al.*, 1994). Finger II from CREA is very similar to finger I from Zif268 across the majority of the helical region (including the -1, +2, +3 an +6 positions), yet finger I from Zif268 recognises G C/T/A G, but Finger II of CREA is able to recognise G/C(6) C/T(5) G sequences. In comparison finger I from CREA is

quite dissimilar to finger I from ADR1 except at the -1, +2, +3 and +6 positions but yet they recognise the same G G/A G triplet. This suggests that while the model is applicable to many proteins some zinc fingers may contact the DNA differently and this may depend on the presence and type of additional or neighbouring fingers.

CHAPTER 5

Chapter 5

Generation of Antibodies Toward CREA

In this chapter the generation of polyclonal antibodies toward two regions of the CREA protein, and use of these in experiments to detect native CREA protein, are discussed. Some preliminary results of binding by nuclear extracts is also presented.

In vivo experiments to determine the molecular mechanism of CREA's involvement in carbon catabolite repression require the detection of CREA in *in vivo* extracts. In order to detect CREA, antibodies were raised to two portions of the protein. There are many questions which can be answered using antibodies, some of which are given below.

The binding studies presented in chapters 3 and 4 were undertaken *in vitro* with an *E. coli* generated fusion protein. In order to study the binding of native CREA from nuclear extracts, it is necessary to be able to identify CREA and antibodies would be an ideal means of doing this. In addition, these experiments may show that other proteins co-precipitate, and if so, antibodies toward CREA would allow an entry to the study of interacting proteins.

The molecular change has been identified in a number of *creA* mutant alleles, but a complete analysis of these mutants requires a study of the protein so formed, for example, its binding, nuclear localisation, stability and protein-protein interaction properties. The generation of specific antibodies will be valuable in this analysis. Another vital question regarding CREA is whether its stability or activity is different in different growth conditions, and once again antibodies that detect CREA will be of importance in these studies. Another question for which antibodies would be useful is the question as to whether CREA functions as a monomer or a multimer, and whether this property is dependent on the growth conditions.

5.1 Raising antibodies against the CREA protein

Antibodies were raised to two fusion proteins. One was the same fusion protein used in chapters 3 and 4, containing the CREA zinc fingers and alanine rich region. Column purified GST-

CREA was used to inoculate a rabbit and positive results using Ochterlony plates indicated that antibodies to the fusion protein had been formed. Since the zinc finger antibodies could be expected to cross hybridise to other zinc finger proteins another portion of the CREA protein was also used to generate antibodies.

Antibodies to the RGR1 similar and acidic regions of the CREA protein were raised by constructing a different fusion protein with GST. The *Eco* RV-*Eco* RV fragment of 470 bp from pCD5 was cloned into the *Eco* RI digested and endfilled pGEX-2T expression vector (Fig 5.1a). *E. coli* cells transformed with the pGEX-*creA*(RGR) construct were harvested and the yield was determined to be 46 mg of total protein/ 100 ml of bacterial culture. The fusion protein was purified on a glutathione sepharose column and SDS gel electrophoresis was used to check that a protein of 46 kDa was produced (Fig. 5.1b). In addition to the expected GST-CREA(RGR) full length fusion protein two smaller proteins ~35 and 30 kDa, and a 70 kDa protein from *E. coli* were visible. This mixture contained GST-CREA(RGR) as the main protein and was used to inoculate two rabbits.

5.2 Use of antibodies in Western blots

Only one rabbit was inoculated with GST-CREA, and thus this serum (ZF serum) was used in further experiments. For GST-CREA(RGR), two rabbits were inoculated and thus the sera were tested to determine which had the highest titre of antibodies (Fig. 5.2a,b). The serum from rabbit 2, injected with GST-CREA(RGR), contained as many antibodies to the fusion protein as rabbit 1 but had fewer antibodies to the 70 kDa protein and other proteins and thus was used for further experiments.

5.2.1 Western blots of *E. coli* expressed protein extracts

The CREA-GST fusion constructs contain thrombin digestion sites at the boundary between GST and the cloned CREA protein. In order to determine whether there was a significant proportion of the polyclonal antibodies directed toward the CREA portion of the fusion protein,

or whether the majority were directed toward GST, Western analysis was carried out on column purified extracts digested with thrombin. For GST-CREA, the two protein fragments produced after digestion with thrombin are approximately 26 kDa in size and could not be distinguished on an SDS gel for Western analyses (Fig. 5.2d), and these bands do not resolve on a higher percentage acrylamide gel (results not shown). The CREA(RGR) portion of GST-CREA(RGR) migrates faster than GST due to its smaller size (20 kDa) and Western analysis shows that the RGR serum contains antibodies toward the CREA portion of this fusion protein (Fig. 5.2c). There is very little decrease in the immunodetection to the uncut fusion protein bands, suggesting that thrombin digestion was partial. The ZF serum was not able to detect the CREA(RGR) peptide. The small peptides present in the track containing GST-CREA digested with thrombin (Track 4, Fig. 5.2d), which may be degradation products of CREA(ZF), were not recognised by the RGR antibodies. Thus the RGR1 serum contained antibodies to the CREA protein at a titre that allowed detection of expressed protein. Presumably the ZF serum contained antibodies to the CREA protein. However, even though the ZF and RGR constructs overlap by 34 amino acids, the antibodies do not significantly cross react to the CREA moieties.

5.2.2 Western blots of *A. nidulans* proteins

5.2.2.1 Strains which lack CREA

In order to determine whether the antibody sera were successful in detecting CREA *in vivo* it was necessary to compare wild type with a null mutant strain. As discussed in section 1.2.3.4, a strain with a deletion of the *creA* locus and its flanking DNA produces spores which germinate but do not grow to colonies. Recently a *creA* mutation, *creA303*, has been characterised and found to contain a stop codon at amino acid position 70 (R.A. Shroff and J.M. Kelly, pers. comm.). This strain is viable yet it lacks a functional zinc finger region. Therefore, *creA303* represents an allele predicted to have no functional CREA, and thus it was used in Western analysis.

5.2.2.2 Western analyses of wild type and *creA303* containing strains

Nuclear extracts from glucose grown cultures were prepared (section 2.2.12.1) and analysed by SDS PAGE for the presence of protein bands. Unless otherwise stated, approximately equal

Figure 5.1a Fusion protein construct between GST and the RGR1 similar and acidic regions of CREA

The *Eco* RV-*Eco* RV fragment from pCD5 was cloned into the *Eco* RI digested and end filled expression vector pGEX-2T to give the plasmid pGEX-*creA*(RGR). Details of the expression vector are given in the legend to Figure 3.1a.

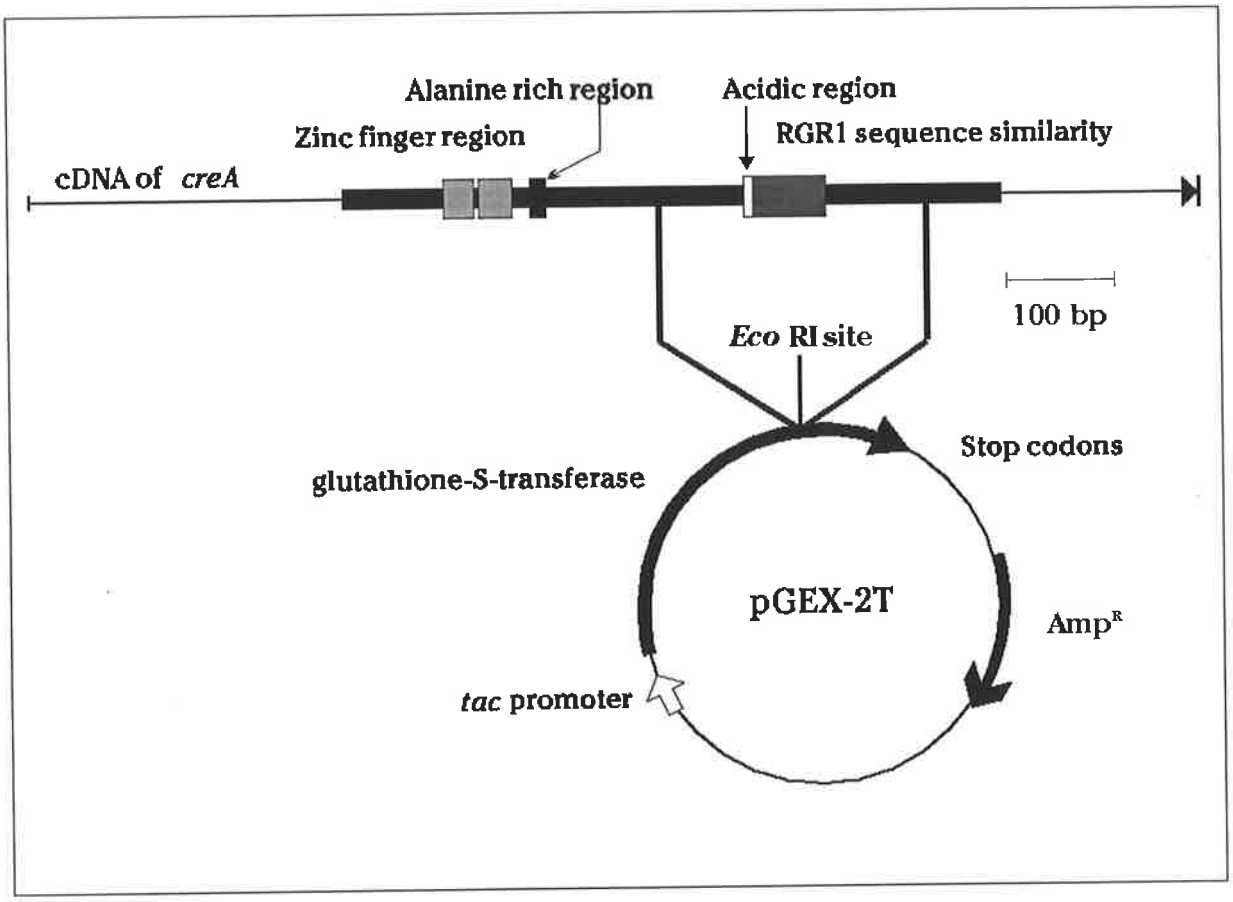
Figure 5.1b Electrophoretic analysis of the GST-CREA(RGR) fusion protein

A 10% SDS gel was loaded with the following *E. coli* extracts and stained with Coomassie Blue®:

Track 1 1 μg prestained molecular weight markers
Track 2 150 μg total protein extract containing GST-CREA
Track 3 13 μg column purified GST-CREA
Track 4 12 μg column purified GST-CREA(RGR)
Track 5 180 μg total protein extract containing GST-CREA(RGR)

The position of the GST-CREA(RGR) fusion protein (lower arrow), the GST-CREA fusion protein (middle arrow) and the 70 kDa contaminant protein (upper arrow) is shown. Approximate molecular weights in kilodaltons are given on the left hand side of each gel.

a)


Alanine rich region
Acidic region
Zinc finger region
RGR1 sequence similarity
cDNA of creA
100 bp
Eco RI site
Stop codons
glutathione-S-transferase
pGEX-2T
Amp^R
tac promoter

b)

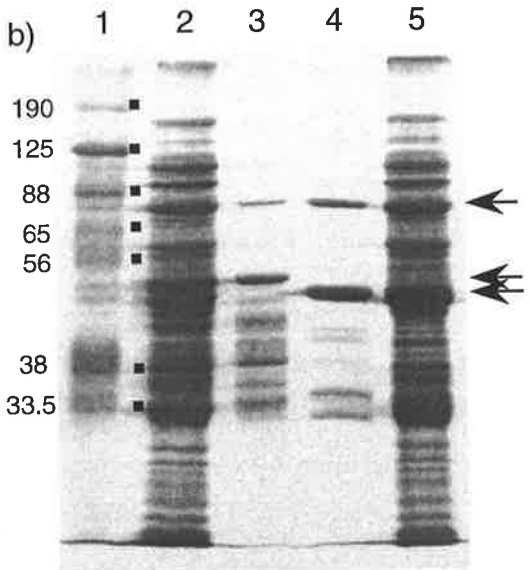

1
2
3
4
5
190
125
88
65
56
38
33.5

Figure 5.2a/ b Comparison of detection using the two anti-GST-CREA(RGR) sera

Two 10% SDS gels were loaded with the following *E. coli* extracts:

Track 1 1 µg prestained molecular weight markers (sizes are shown in kDa on the left)
Track 2 75 µg total protein extract containing GST
Track 3 15 µg total protein extract containing GST
Track 4 1.5 µg total protein extract containing GST
Track 5 0.15 µg total protein extract containing GST
Track 6 115 µg total protein extract containing GST-CREA(RGR)
Track 7 23 µg total protein extract containing GST-CREA(RGR)
Track 8 2.3 µg total protein extract containing GST-CREA(RGR)
Track 9 0.23 µg total protein extract containing GST-CREA(RGR)
Track 10 2 µg column purified GST-CREA(RGR)

after electrophoresis gels were Western transferred using a semi-dry blotting apparatus at 40 mA for 0.5 h. Immunodetection (section 2.2.10.2) was carried out with 50 µl rabbit 1 (RGR) serum (Fig. 5.2a) or rabbit 2 (RGR) serum (Fig. 5.2b). In each case the upper arrow marks the position of the 70 kDa protein and the lower arrow the position of the 46 kDa fusion protein.

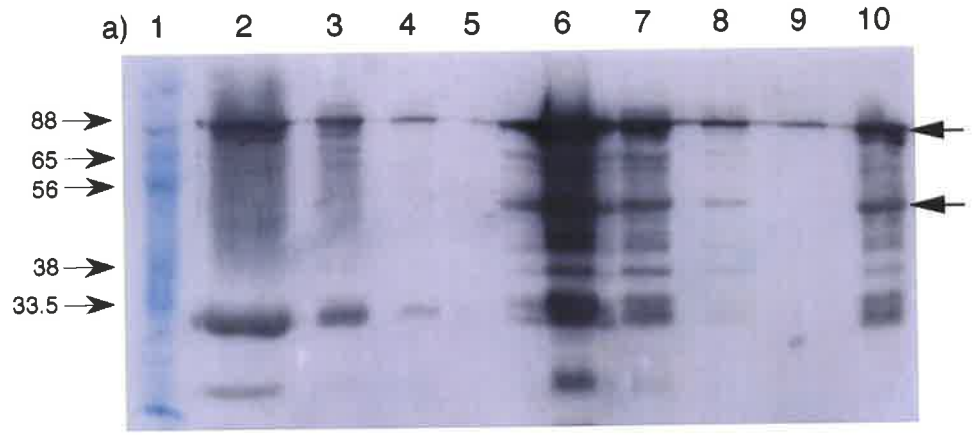
Figure 5.2c/ d Western analysis of the fusion proteins digested with thrombin

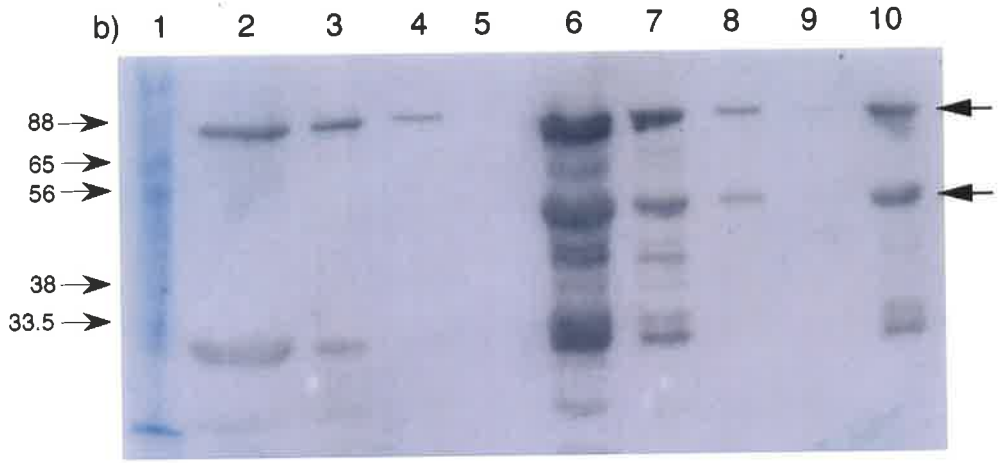
Three 10% SDS gels were loaded with the following total *E. coli* protein extracts:

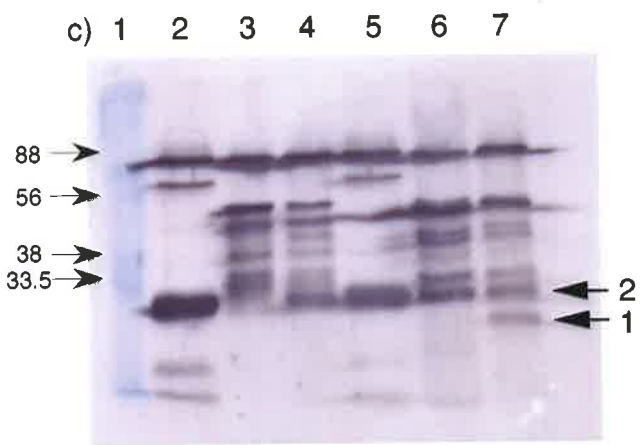
Track 1 1 µg prestained molecular weight markers (sizes are shown in kDa on the left)
Track 2 30 µg total protein extract containing GST
Track 3 3 µg column purified GST-CREA
Track 4 3 µg column purified GST-CREA digested with thrombin
Track 5 30 µg total protein extract containing GST
Track 6 3 µg column purified GST-CREA(RGR)
Track 7 3 µg column purified GST-CREA(RGR) digested with thrombin

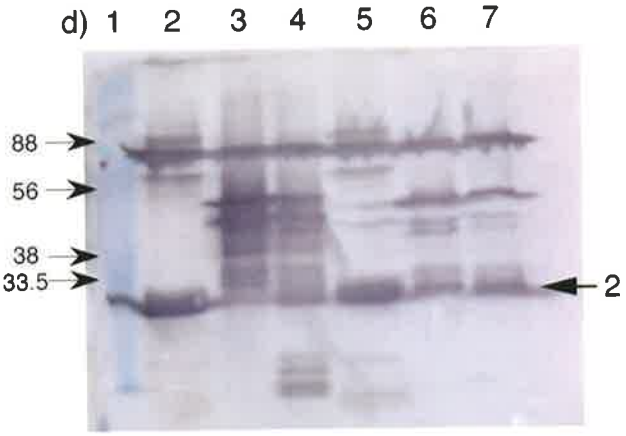
Thrombin digests were carried out at 37°C for 60 min with 100 ng of thrombin (Sigma) per 50 µg of target protein.

After electrophoresis one gel was stained with Coomassie Blue® to confirm equal loadings (not shown) and the other two were transferred using a Trans-Blot (submerged) apparatus (section 2.2.10.1). Immunodetection was carried out with 500 µl RGR serum (Fig. 5.3c) or 500 µl ZF serum (Fig. 5.3d). The presence of the RGR(←1) and GST(←2) portions of the GST-CREA(RGR) fusion protein is shown by arrows.


a) 1 2 3 4 5 6 7 8 9 10
88
65
56
38
33.5


b) 1 2 3 4 5 6 7 8 9 10
88
65
56
38
33.5


c) 1 2 3 4 5 6 7
88
56
38
33.5
2
1


d) 1 2 3 4 5 6 7
88
56
38
33.5
2

amounts of protein as determined by Coomassie Blue® staining were loaded onto gels. Western analysis of nuclear extracts from wild type and a strain containing the *creA303* allele is shown in Figure 5.3a and b. Immunodetection with RGR serum results in a band at 56-65 kDa (arrow number 1 in Fig. 5.3a), which is only slightly larger than the expected size for the native CREA protein. There was no clear band at 56-65 kDa in the track loaded with nuclear extract from a strain carrying the *creA303* allele, however the presence of a band can not be totally ruled out due to the large amount of staining of this extract. The extract from the wild type strain contains other bands at 45-50 and 65-70 kDa (arrow numbers 2 and 3 in Fig. 5.3a). Although identical procedures were used, the patterns of detection from nuclear extracts made at different times was found to vary. For both protein extracts, bands were detected which correspond to much larger proteins. Immunodetection with the ZF antibodies (Fig. 5.3b) produced a band at 45-50 kDa (arrow number 4 in Fig. 5.3b) which was also present in extracts made from a strain containing the *creA303* allele and is thus not CREA. Other bands at 56-65 kDa and 65-70 kDa (arrow numbers 5 and 6 in Fig. 5.3b) were not detected in extracts from a strain containing the *creA303* allele. Thus consistent with the result from the RGR serum, the 56-65 kDa band may be CREA. There are three possible start methionines within the theoretically translated amino acid sequence of CREA. Assuming the largest open reading frame, a size of 52 kDa would be expected, but the protein may have a larger apparent molecular weight due to post-translational modifications. The MIG1 protein was found to be phosphorylated *in vivo* (Treitel and Carlson, 1995) and glycosylation of mammalian transcription factors, such as Sp1, suggests glycosylation may be important in transcription initiation by RNA polymerase II in higher eukaryotes (Jackson and Tjian, 1988).

The overall amount of immunodetection using the RGR serum was higher in extracts from a *creA303* containing strain compared to the wild type. This was not expected. Since *A. nidulans* is likely to contain a large number of zinc finger containing proteins, it is surprising that the ZF serum appears to cross react less than the RGR serum.

Figure 5.3a/ b Western analysis of nuclear extracts from "wild type" and *creA303* containing strains

Two 10% SDS gels were loaded with the following nuclear extracts made from mycelia grown in glucose media:

Track 1 prestained molecular weight markers (sizes are shown in kDa on the left)

Track 2 10 µl "wild type" extract preparation A

Track 3 10 µl "wild type" extract preparation B

Track 4 7.5 µl *creA303*

Nuclear extract preparations A and B were made according to the same procedure (section 2.2.12.1) but using mycelia grown on different days. Samples were electrophoresed and transferred (submerged). Membranes were immunoblotted with either 200 µl RGR serum (Fig. 5.3a) or 300 µl ZF serum (Fig. 5.3b).

The positions of the bands discussed in the text are shown with arrows.

Figure 5.3c Western analysis of total protein extracts from "wild type" and *creA303* containing strains

A 10% SDS gel was loaded with 10 µl of the following total protein extracts:

Tracks 1 and 6 prestained molecular weight markers (sizes are shown in kDa on the left)

Tracks 2 and 7 "wild type" grown overnight in low sulfate media (contains glucose as a carbon source)

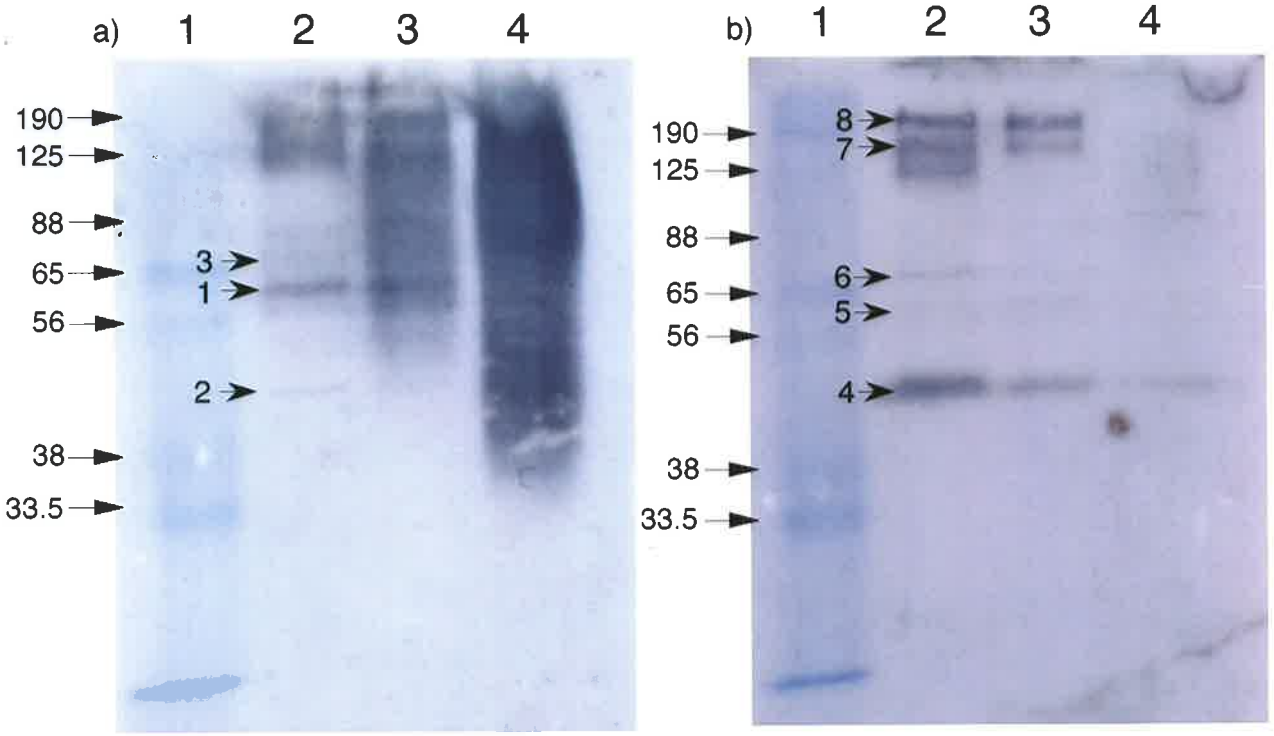
Tracks 3 and 8 *creA303* grown overnight in low sulfate media

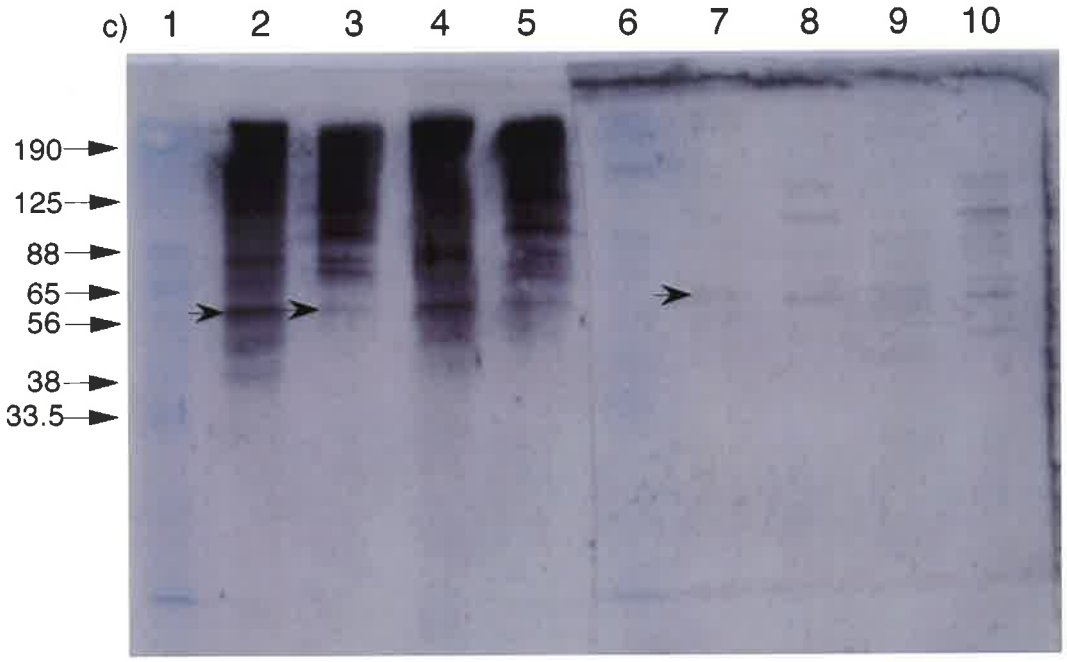
Tracks 4 and 9 "wild type" grown overnight in low sulfate media and then transferred for 1 hr to fresh media

Tracks 5 and 10 *creA303* grown overnight in low sulfate media and then transferred for 1 hr to fresh media

After transfer (submerged) the membrane was cut in half and Immunoblotted with either 200 µl RGR serum (**Tracks 1-5**) or 300 µl ZF serum (**Tracks 6-10**).

The position of the band discussed in the text is shown with arrows.


a)
1
2
3
4
190
125
88
65
56
38
33.5
3
1
2
b)
1
2
3
4
190
125
88
65
56
38
33.5
8
7
6
5
4


c)
1
2
3
4
5
6
7
8
9
10
190
125
88
65
56
38
33.5

There was strong detection with both antibodies to very high molecular weight proteins. Two bands detected with the ZF antibodies were present in wild type extracts (arrow numbers 7 and 8 in Fig. 5.3b) but not in the *creA303* extract. It is possible that these correspond to CREA.

The procedure used to make nuclear extracts was long and complex, and if CREA is an unstable protein, it may have been degraded. Therefore *Aspergillus* total protein extracts were used in Western analyses. An experiment was carried out using total *Aspergillus* protein extracts from mycelia grown in glucose containing but low sulfate media. In order to determine whether ZF and RGR antibodies were detecting a protein of the same size, duplicate wild type and *creA303* samples were electrophoresed on the same gel. Immunodetection (Tracks 2,3 and 7,8, Fig. 5.3c) shows that many very high molecular weight proteins were detected with the RGR serum and the overall level of detection was greater with the RGR serum, compared to the ZF serum. One band at approximately 56-65 kDa was detected in both wild type and *creA303* containing strains with the RGR serum (albeit considerably more strongly in the wild type strain) and only in the wild type strain with the ZF serum (arrowed). If this band is CREA, then the RGR antibodies must be cross reacting with another protein of approximately the same size, whereas the ZF antibodies were not. This is consistent with the relative intensities of the bands detected with the RGR serum. This experiment was repeated after transferring mycelia to fresh media for one hour (Fig. 5.3c, Tracks 4,5 and 9, 10) because Strauss *et al.*, (1995a) reported that *creA* mRNA increases transiently 10 minutes after transferring mycelia to fresh culture medium. One explanation for this observation may be that since overnight growth results in a significant drop in the pH of the culture medium, transferring to fresh liquid media will result in a pH shift which may induce a transient increase in *creA* mRNA expression. The pattern of immunodetection was similar before and after transfer to fresh media for one hour, except that in the *creA303* strain an additional band was present after transfer that was larger than that identified as potentially CREA. The 56-65 kDa band did not appear to increase in intensity in the extracts from transferred wild type.

Although the RGR antibodies were expected to be more specific for detecting CREA, in reality they were able to detect more proteins than the ZF serum. In addition extracts from the *creA303* containing strain, which is predicted to lack the CREA protein, produced a high level of overall detection with the RGR serum. Therefore it may be that the ZF serum is more specific for the CREA protein compared to the RGR serum. The experiments presented in Figure 5.3 are consistent with the suggestion that a band in the size range 56-65 kDa may be CREA if the RGR serum is assumed to cross-react.

5.2.2.3 Transformants containing multiple copies of the *creA* gene

It was not known whether the amount of CREA protein in wild type or mutant strains was sufficient for immunodetection and therefore strains were constructed that contained multiple copies of the *cre*A gene. *A. nidulans* strain C43 (Table 2.1) was co-transformed with pMOO6 and pANC6 to arginine prototrophy. Transformants were either overlayed with protoplast media plus 50 mM allyl alcohol (P2, 11 and transformants designated CO) or replica plated to 1% sucrose plus 50 mM allyl alcohol (designated CR transformants). Transformants were screened in this way to detect tighter control of the alcohol dehydrogenase locus by multiple copies of the *creA* gene.

40 transformants were chosen for further analysis. To determine whether these transformants contained tandemly duplicated sequences or whether the extra copies had been inserted elsewhere in the genome Southern analysis was carried out using DNA samples prepared from each of these transformants. Two restriction endonucleases were chosen, one, *Bam* HI, cuts pANC6 once and the other, *Xho* I, does not cut within the pANC6 sequence. The recipient strain C43 contains the *creA204* point mutation (R.A. Shroff and J.M. Kelly, submitted, Table 5.1) and therefore produces a wild type restriction pattern upon hybridisation with the insert of pANC6 (Fig. 5.4a). Many transformants had a high copy number of *creA* genes as shown by the increase in hybridisation to the insert of pANC6, including transformant 11 (result not shown). Digests with *Xho* I showed that transformants CR1, P2 and CR6 contain tandemly repeated pANC6 sequences (Fig. 5.4a,b).

Ten transformants were chosen for Northern analysis using total RNA from each strain. Northern analysis of seven of these strains is shown and confirms the presence of full length mRNA in greater amounts in five of the transformants (Fig. 5.4c.). It appears that there are two RNA bands hybridising with the probe in the transformants, however the lack of hybridisation in the region between the two bands may be due to the presence of rRNA which reduces the transfer efficiency of other RNA species since it is in excess. From these results transformants CR1, CR6 and 11 were chosen for protein analysis due to their high expression of *creA* mRNA. CR1 and CR6 grew well, indicating that essential genes had not been disrupted by transforming sequences. Transformant 11 showed impaired growth.

5.2.2.4 Western analyses of nuclear extracts of *A. nidulans*

Nuclear extracts of mycelia grown in glucose containing media were made from wild type, strains containing mutant *creA* alleles (Table 5.1) and from strains containing multiple copies of the *creA* gene. The theoretical size expected for the CREA proteins encoded by genes with mutations in *creA* (assuming the longest open reading frame) is shown in Table 5.1. These extracts were used in Western analyses and the results are shown in Figure 5.5a and b. Using either serum, the pattern of immunodetection for the two multiple copy transformants of *creA* was similar to the wild type, whereas the mutant strains shown display quite different patterns of detection. This reflects the derepressed nature of the mutant strains. Track 5, containing an extract from a strain carrying the *creA220* allele was relatively under loaded, as determined by staining with Coomassie Blue®. In the wild type strain the major band detected by the RGR antibodies was in the range 65-70 kDa (Fig. 5.5a) which is larger than the significant band detected previously (arrow number 1 in Fig. 5.3a) at 56-65 kDa. A minor band at 45-55 kDa band was also present (Fig. 5.5a). However both of these are present in all the mutant strains shown including those that are expected to produce a truncated CREA protein. Thus from the immunodetections using the RGR serum there was no clear candidate band for CREA seen in this experiment. The major band detected with the ZF antibodies was at 45-50 kDa, but it was present in extracts from all the strains shown (Fig. 5.5b). A 56-65 kDa minor band is visible

Table 5.1 Molecular basis of mutations within the *creA* locus✪

Allele	Mutation	Theoretical size of protein	Change to amino acid sequence †
creA1	G→A at bp# 264	52 kDa	G→D at aa# 88
creA30	pericentric inversion - breakpoint at bp# 391	19 kDa	wild type until aa#130 + RDGAALVWLGPLGLLGSIR*
creA204	G→T at bp# 320	52 kDa	D→Y at aa# 107
creA218	ΔG at bp# 731	32 kDa	wild type until aa# 244 + DFLLFQPTPSPTA*
creA220	C→A at bp# 813	34 kDa	S→ termination codon at aa# 271
creA221	7 bp insertion 'GCGTCAT' between bp# 808-809	35 kDa	wild type until aa# 271 + CVTSRQAFKT*
creA225	G→A at bp# 320	52 kDa	D→ N at aa# 107
creA303	G→T at bp# 208	8 kDa	E→ termination codon at aa# 70
creA304	four bp deletion from # 337-340 'TCGC'	19 kDa	wild type until aa# 112+ESITTPTQDVETRLNTWR QPPQLQLRTKMVARWRTTLDQ*

✪ numbering based on the sequence published in Dowzer and Kelly, (1991) and data is from Shroff *et al.*, (submitted) and R.A. Shroff and J.M. Kelly (pers. comm.)

† aa = amino acid, * = stop codon

but this was also present in the truncation mutant strains *creA30*[1] (although at a much reduced level) and *creA218* and thus could only be CREA if the serum was cross-reacting with another protein of apparently the same size. There is a small band present in the *creA218* extract which was detected by both the RGR and ZF sera, but it is too small to be full length CREA, which is predicted to be 32 kDa in this strain. Therefore, the antibodies raised against the RGR and ZF portions of CREA in this study were not able to reliably detect the native CREA protein in nuclear extracts.

To investigate the proteins that were in the 65-70 kDa size range, a set of gels were electrophoresed for a longer period and these are shown in Figure 5.5c and d. Immunodetections with the RGR antibodies show that the significant band at 65-70 kDa was also present at a much reduced level in the *creA303* strain. Furthermore, *creA220* also contains a protein in this region. There is a faint band detected in the wild type, *creA204* and multiple copy transformant strain which is not present in the truncation mutant strains (arrowed). In the blot immunodetected with ZF antibodies a band in approximately the same location is visible (arrowed). If these represent CREA, then the CREA molecules must have been post translationally modified. The very high molecular weight bands detected previously in the wild type strain (Fig. 5.3a and b) were not present in these blots, except in the multiple copy transformant, CR1.

Due to the lack of consistency between experiments, it is not possible to say which, if any, of the bands detected in Western experiments are CREA .

5.2.2.5 Western blots of *A. nidulans* total protein extracts

Western analyses of total protein extracts from two growth conditions, 1% glucose and 1% arabinose, are shown in Figure 5.6. The wild type strain has higher levels of *creA* message in arabinose and glycerol compared to glucose grown cultures (Dowzer, 1991). Thus total

[1]The *creA30* mutation was selected because it conferred resistance in a *frA*-1 strain to the toxicity of 1% D-mannitol. The mechanism by which *creA30* confers this resistance is not understood (Arst *et al.*, 1990).

Figure 5.4a/ b Southern analysis of transformants containing multiple copies of the *creA* gene

1 μg of digested genomic DNA from the recipient strain, C43 (containing the *creA204* allele) and from various strains co-transformed with pMOO6 and pANC6, was digested with either *Bam* HI (B) or *Xho* I (X) and electrophoresed on a 0.7% agarose gel. After transfer, filters were hybridised with the insert of pANC6. λ= 200 ng labelled λ/ *Hind* III (sizes are in kb).

w=wells

Figure 5.4c Northern analysis of transformants containing multiple copies of the *creA* gene

and from the recipient strain (C43)

20 μg of total RNA from transformants containing multiple copies of the *creA* gene was electrophoresed on a denaturing agarose gel, transferred to Zetaprobe® and hybridised with the insert of pANC6. As judged by ethidium bromide staining the loadings in the tracks are approximately equal. An autoradiograph of the Northern hybridisation is shown.

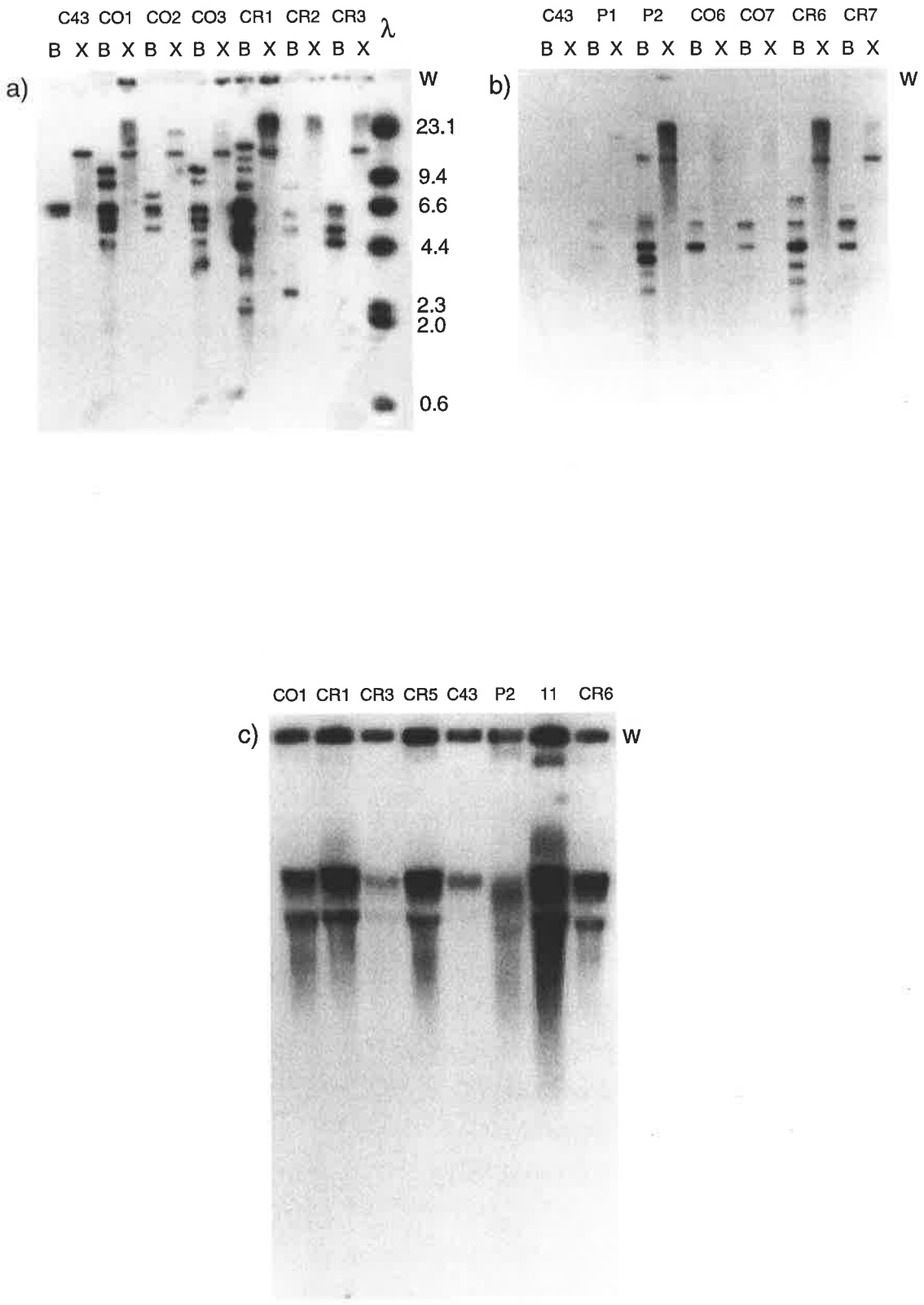

a)
C43 CO1 CO2 CO3 CR1 CR2 CR3 λ
B X B X B X B X B X B X B X
W
23.1
9.4
6.6
4.4
2.3
2.0
0.6
b)
C43 P1 P2 CO6 CO7 CR6 CR7
B X B X B X B X B X B X B X
W
c)
CO1 CR1 CR3 CR5 C43 P2 11 CR6
W

Figure 5.5 Western analysis of nuclear extracts

Figure 5.5a and b

Two 10% SDS gels loaded with 5 µl of the following nuclear extracts from glucose grown cultures:

Track 1 prestained molecular weight markers (sizes are shown in kDa on the left)
Track 2 "wild type" extract preparation F
Track 3 no extract
Track 4 *creA204* extract preparation C
Track 5 *creA220*
Track 6 *creA218*
Track 7 *creA30* extract preparation D
Track 8 multiple copy transformant CR1
Track 9 multiple copy transformant CR6

were electrophoresed and transferred (submerged). Membranes were Immunoblotted with either 900 µl RGR serum (Fig. 5.5a) or 600 µl ZF serum (Fig. 5.5b).

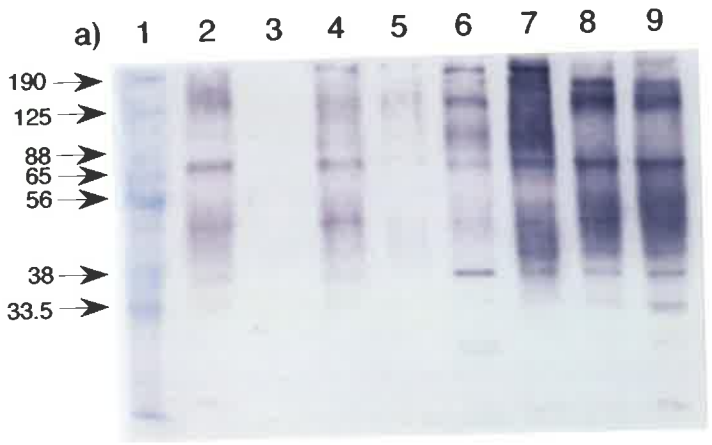
Figure 5.5c/ d

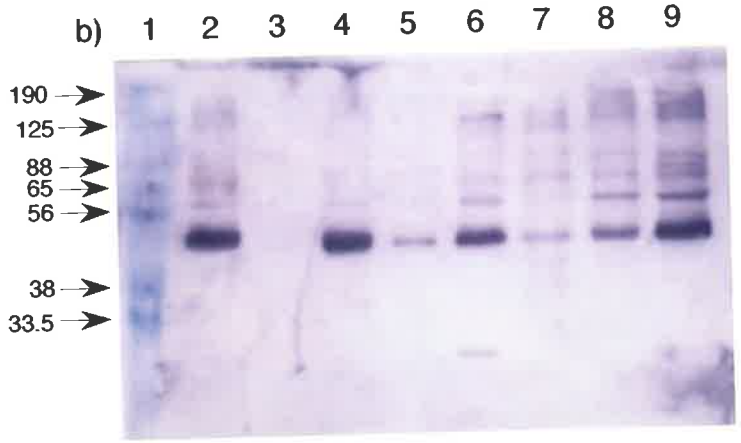
Two 10% SDS gels loaded with 5 µl of the following nuclear extracts from glucose grown cultures:

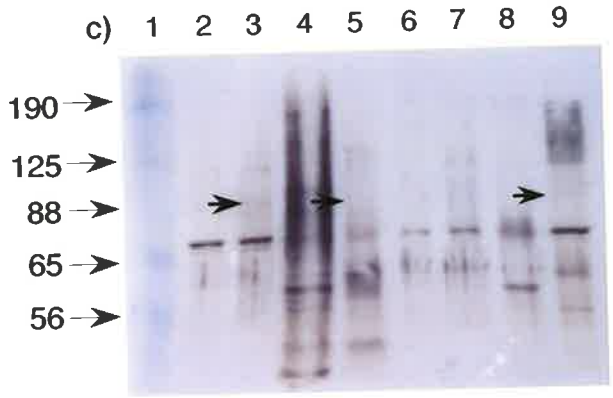
Track 1 prestained molecular weight markers (sizes are shown in kDa on the left)
Track 2 "wild type" extract preparation E
Track 3 "wild type" extract preparation D
Track 4 *creA303*
Track 5 *creA204* extract preparation A
Track 6 *creA204* extract preparation B
Track 7 *creA220* extract preparation A
Track 8 *creA220* extract preparation B
Track 9 multiple copy transformant CR1

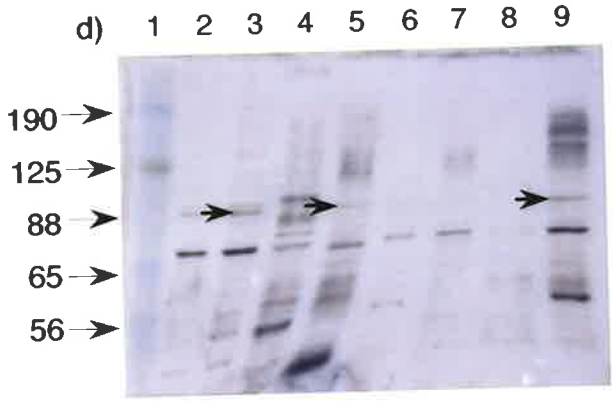
were electrophoresed and transferred (submerged). Membranes were Immunoblotted with either 900 µl RGR serum (Fig. 5.5c) or 600 µl ZF serum (Fig. 5.5d).

The position of the band discussed in the text is shown with arrows.


a)
1 2 3 4 5 6 7 8 9
190
125
88
65
56
38
33.5


b)
1 2 3 4 5 6 7 8 9
190
125
88
65
56
38
33.5


c)
1 2 3 4 5 6 7 8 9
190
125
88
65
56


d)
1 2 3 4 5 6 7 8 9
190
125
88
65
56

protein extracts from arabinose growth conditions were used in Western analyses on the assumption that the higher levels of mRNA may result in higher levels of CREA protein. The number of bands detected in total extracts by both sera was very large (Fig. 5.6). The major band detected with the RGR serum was 45-55 kDa in size, but it was present in all strains. A protein of similar size was detected with the ZF serum (arrowed) which was present at a reduced level in arabinose grown cultures but present in all strains. A larger protein (60-65 kDa) was detected with the RGR serum which was present in wild type and the *creA204* strain grown in glucose media but not present to the same level when grown in arabinose media (arrowed). This band, however, was present at a low level in the truncation mutants *creA30*, *creA221*, *creA303* and *creA304* meaning that this can only be CREA if the RGR serum was cross-reacting to another protein of the same size. Immunodetections with the ZF antibodies did not indicate any bands which could be CREA.

Attempts were made to remove antibodies which were not directed at the CREA protein in order to reduce false detections. One attempt was to remove the anti-GST antibodies from the ZF serum using a column containing GST. The eluate from this column was passed through a column containing the GST-CREA fusion protein and antibodies removed using acidic (100 mM glycine, pH 3) and basic (100 mM triethylamine, pH 11) conditions. This procedure did not produce antibodies of a high enough titre to detect proteins in Western blots or immunoprecipitations (results not shown). Using the RGR serum, an attempt was made to purify antibodies directed only against the CREA portion of the fusion protein, by using strips of membrane from a Western transfer which contained the CREA peptide from the thrombin digested GST-CREA(RGR) fusion protein. No antibodies could be removed from these strips with 100 mM glycine (pH 3) or 100 mM triethylamine (pH 11). Another attempt to remove the anti-GST antibodies consisted of pre-incubating the RGR serum with GST total protein extract. A volume 15 times the volume of the RGR serum was necessary to remove all the antibodies toward the GST peptide. Preincubated RGR serum showed reduced overall detection in Western analysis of nuclear extracts (consistent with a low titre of antibodies toward CREA), but there was no band consistent with its being CREA.

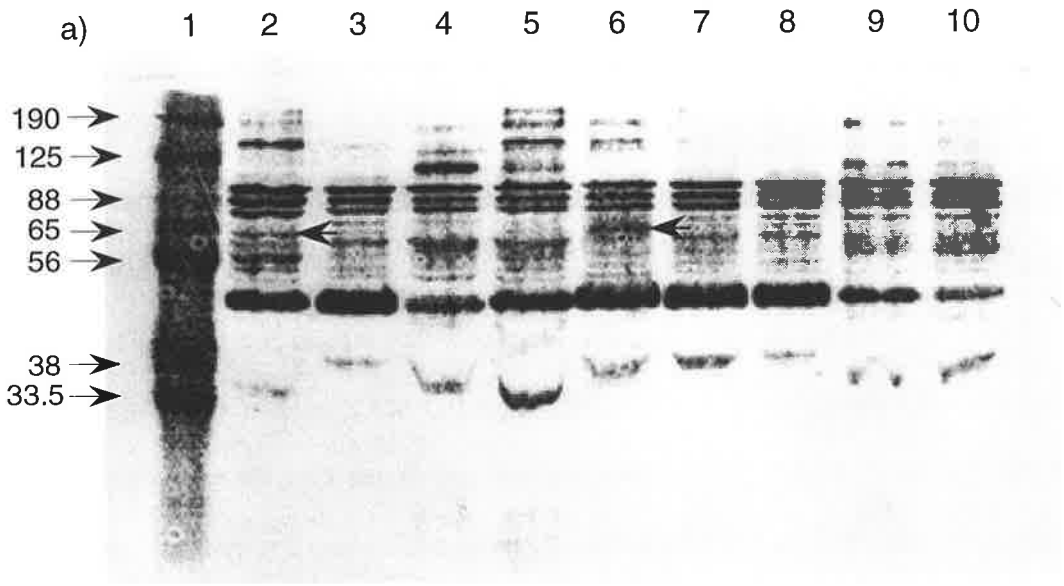
Figure 5.6 Western analysis of *A. nidulans* total extracts from cultures grown in glucose and arabinose growth conditions

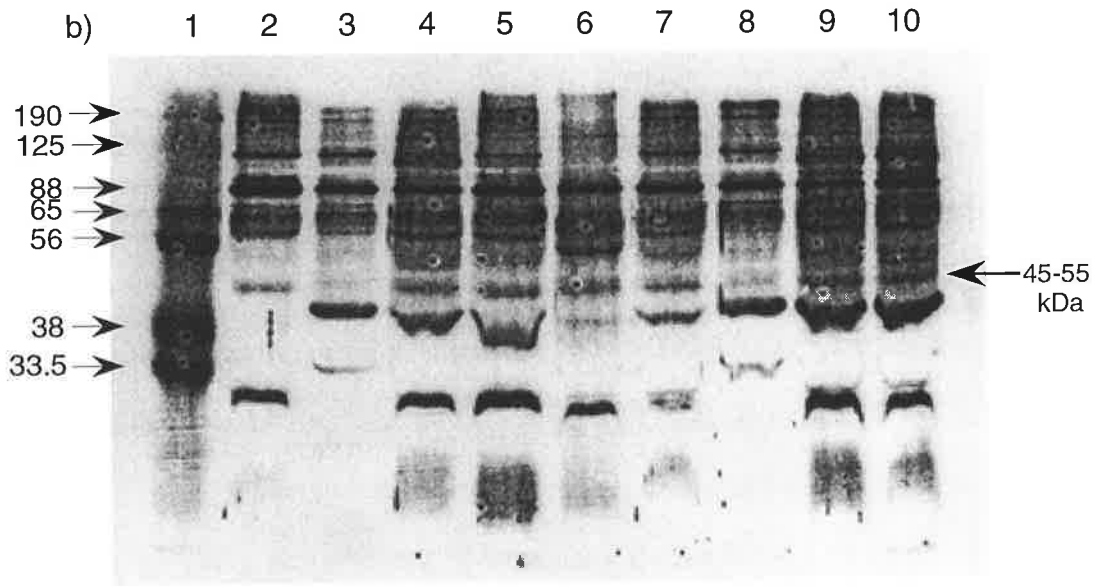
Two 10% SDS gels were loaded with 5 µl of the following total protein extracts:

Track 1 prestained molecular weight markers (sizes are shown in kDa on the left)
Track 2 "wild type" grown in 1% glucose media
Track 3 "wild type" grown in 1% arabinose media
Track 4 *creA30* grown in 1% arabinose media
Track 5 *creA221* grown in 1% arabinose media
Track 6 *creA204* grown in 1% glucose media
Track 7 *creA204* grown in 1% arabinose media
Track 8 "wild type" grown in 1% arabinose media
Track 9 *creA303* grown in 1% arabinose media
Track 10 *creA304* grown in 1% arabinose media

were electrophoresed and transferred (submerged). Membranes were Immunoblotted with either 900 µl RGR serum (Fig. 5.6a) or 600 µl ZF serum (Fig. 5.6b).

The positions of the bands discussed in the text are shown with arrows.


a)
1 2 3 4 5 6 7 8 9 10
190
125
88
65
56
38
33.5


b)
1 2 3 4 5 6 7 8 9 10
190
125
88
65
56
38
33.5
45-55
kDa

5.3 Mobility shift assays using *A. nidulans* extracts

Mobility shift assays were undertaken using nuclear extracts from *A. nidulans* to determine whether binding by native CREA protein could be detected. Nuclear extracts from wild type and *creA30* strains were used in gel mobility shift assays on the ALU148 fragment from *amdS* (Fig. 5.7a). Protein(s) in nuclear extracts were shown to bind this fragment and form a very low mobility complex. Binding was not detected in wild type and *creA30* extracts that had been frozen and thawed twice as the preparation "A" extracts had been. Binding by the extract from the *creA30* containing strain which had been frozen and thawed just once (preparation "B") was slightly reduced compared to wild type, but this varied between experiments (see Fig. 5.7b). Pre-incubation of these extracts with ZF antibodies did not appear to affect binding. In the experiment shown in Figure 5.7b, there was an indication that extracts from a *creA30* containing strain produced a retarded complex that was slightly smaller than the complex in the wild type strain. This is what would be expected since the *creA30* allele encodes a truncated form of CREA but one that contains the DNA binding domain. The ALU148 fragment contains sequences that are binding sites for other proteins including AMDA, FACB and possibly AREA, and thus a short oligonucleotide containing CREA binding sites was used.

A comparison of binding by various extracts with the ALU148 fragment and the ONC1/ONC7 oligonucleotide is shown in Figure 5.7c. Once again binding by the extract from a wild type strain to the ALU148 fragment was seen. The extract from the *creA220* strain was able to bind ALU148 producing a faster migratory complex compared to the retarded complex of the wild type strain. This is in keeping with the molecular data of the mutation that the zinc finger region was intact in this mutant but that a truncated protein of 34 kDa would be formed (Table 5.1). The lack of binding by the *creA30* extract was probably due to its age. A nuclear extract from glycerol grown wild type (derepressing conditions) was also able to bind to the ALU148 fragment and produced a retarded complex that was even slower in migration than the retarded complex formed by the wild type strain grown in glucose growth conditions. A comparison with binding to the ONC1/ONC7 oligonucleotide shows that similar results are found (see

arrows in Tracks 7-10 of Fig. 5.7c) which suggests that retardation of the ALU148 fragment was due, at least in part, to CREA.

CREA was not able to bind oligonucleotides 31/32 *in vitro,* however, since this sequence was protected in DNase I footprinting it may be important *in vivo*. For this reason oligonucleotide 31/32 was used in a gel mobility shift assay (Fig. 5.7d). As shown there was binding to this oligonucleotide by a protein present in wild type and *creA30* extracts. This complex did not change in mobility in the *creA30* extract containing tracks as the CREA protein would be expected to do and therefore, is not due to CREA.

A range of binding experiments were performed using total protein extracts from *A. nidulans* and the ONC1/ONC7 oligonucleotide. The results varied greatly between repeated extracts, and no conclusions could be drawn.

5.4 Immunoprecipitations

Antibody sera that are unable to produce satisfactory immunodetections in Western analyses are sometimes able to give good results when used in immunoprecipitations (Harlow and Lane, 1988). The ZF serum was used in preliminary immunoprecipitations of wild type, multiple copy transformant 11 and *creA30* strains grown overnight in low sulfate media with added $^{35}SO_4^{2-}$ (section 2.2.14). The results show that proteins were immunoprecipitated but that the pattern was very similar in all three strains and more than one band was produced (Fig. 5.8).

The RGR serum was used in immunoprecipitations of labelled wild type and *creA303* total protein extracts, however, again many proteins were immunoprecipitated with these antibodies. Comparisons of the proteins immunoprecipitated with the RGR and ZF sera gave inconclusive results (not shown).

Figure 5.7 Gel mobility shift assays of *A. nidulans* nuclear extracts

Figure 5.7a
Nuclear extracts from mycelia grown in glucose media were incubated with the labelled ALU148 fragment (from *amdS*) for 15 min and loaded on a mobility shift gel:

Track 1 binding buffer
Track 2 5 µl "wild type" preparation A
Track 3 20 µl *creA30* preparation A
Track 4 10 µl "wild type" preparation B
Track 5 10 µl *creA30* preparation B
Track 6 10 µl "wild type" preparation B pre-incubated for 15 mins with 2 µl ZF serum
Track 7 10 µl *creA30* preparation B pre-incubated for 15 mins with 2 µl ZF serum
Track 8 5 µl "wild type" preparation A pre-incubated for 15 mins with 2 µl ZF serum
Track 9 20 µl *creA30* preparation A pre-incubated for 15 mins with 2 µl ZF serum

note: preparation "A" had been frozen and thawed twice.

Figure 5.7b
Nuclear extracts from mycelia grown in glucose media were incubated with the labelled ALU148 fragment (from *amdS*) for 15 min and loaded on a mobility shift gel:

Track 1 binding buffer (no added extract)
Track 2 10 µl "wild type" preparation C
Track 3 10 µl *creA30* preparation C
Track 4 10 µl "wild type" preparation C pre-incubated with 4 µl ZF serum
Track 5 10 µl *creA30* preparation C pre-incubated with 4 µl ZF serum

Figure 5.7c
Labelled ALU148 fragment (**Tracks 1-5**) or ONC1/ONC7 oligonucleotide (**Tracks 6-10**) was incubated with 10 µl of nuclear extract and electrophoresed on a mobility shift gel.

Track 1 binding buffer (no added extract)
Track 2 "wild type" extract preparation D
Track 3 *creA220* extract preparation B
Track 4 *creA30* preparation B[1]
Track 5 "wild type" grown with glycerol as the carbon source
Track 6 binding buffer (no added extract)
Track 7 "wild type" extract preparation D
Track 8 *creA220* extract preparation B
Track 9 *creA30* preparation B[1]
Track 10 "wild type" grown with glycerol as the carbon source

Figure 5.7d
A mobility shift assay of the oligonucleotide 31/32 incubated with 2-10 µl of nuclear extract made from glucose grown mycelia. **Track 1** contains labelled 31/32 oligonucleotide only. **Tracks 2-4** contain extract from a "wild type" strain and **Tracks 5-7** contain extract from a *creA30* containing strain.

[1]This was the same extract as was used in the experiment shown in Figure 5.7a except that the protein extract had been stored at -80 for the intervening 2 years.

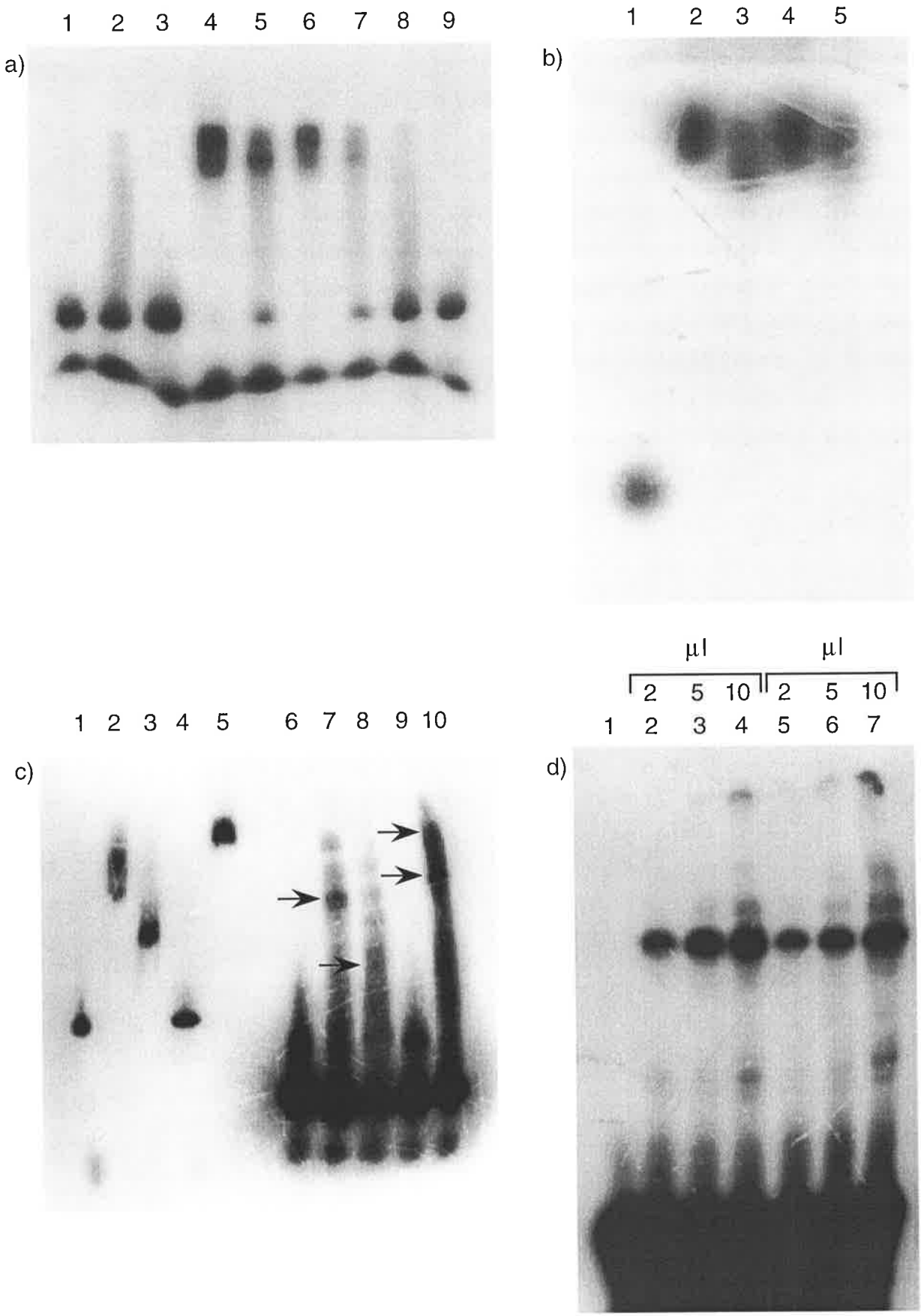
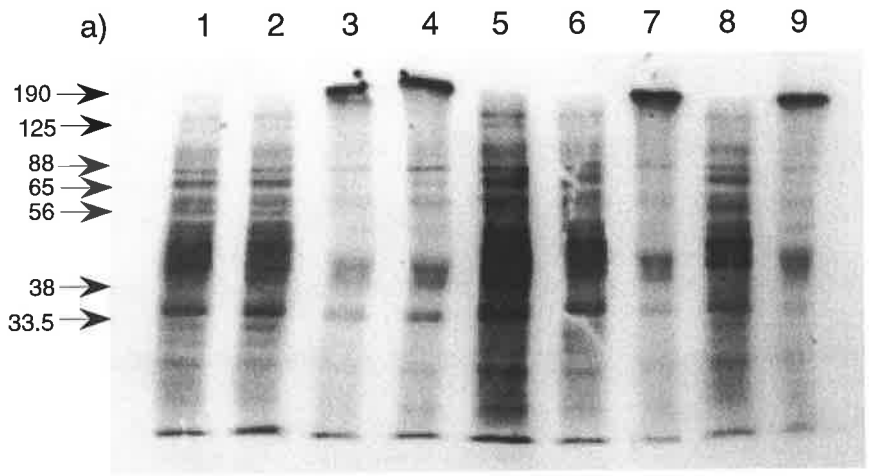

a)
1 2 3 4 5 6 7 8 9
b)
1 2 3 4 5
c)
1 2 3 4 5 6 7 8 9 10
d)
µl µl
2 5 10 2 5 10
1 2 3 4 5 6 7

Figure 5.8 Immunoprecipitation of *A. nidulans* total protein extracts with the ZF serum

in vivo labelled total protein extracts were immunoprecipitated with 15 µl of ZF serum. A 10% SDS PAG was loaded with:

Track 1 "wild type" total protein extract
Track 2 "wild type" supernatant from the precipitation
Track 3 "wild type" immunoprecipitated
Track 4 "wild type" immunoprecipitated
Track 5 11 multiple copy transformant total protein extract
Track 6 11 multiple copy transformant supernatant from the precipitation
Track 7 11 multiple copy transformant immunoprecipitated
Track 8 *creA30* total protein extract
Track 9 *creA30* immunoprecipitated

Shown is an autoradiograph. Approximate sizes are shown (in kDa) on the left hand side of the gel.


a)
1
2
3
4
5
6
7
8
9
190
125
88
65
56
38
33.5

5.5 Discussion

The antibodies generated to the ZF and RGR domains of the protein were able to detect CREA in *E. coli* expressed fusion proteins, but their specificity was insufficient for use in Western analysis, either of total protein or nuclear extracts. Mobility shift assays were undertaken using the ALU148 fragment, oligonucleotides and *in vivo* extracts of CREA. Retardation was shown but the interpretation of the results was difficult without specific antibodies to prove that retarded bands seen were due to CREA. These experiments were also complicated by the fact that repeated extracts that were, as far as is experimentally possible, grown under the same conditions, gave varied results. Treatment of the *A. nidulans* extracts with the antibody preparations gave no clear evidence of super retardation, probably due to the very low titre of antibodies directed at CREA. These experiments have shown that the antibodies produced against CREA not only react with CREA but also cross react with many other *A. nidulans* proteins, some of which may be related to glutathione-S-transferase. Preliminary work using the obvious approach of removing cross reacting antibodies by chromatography through GST containing columns and preincubation with extracts of an *A. nidulans creA* deleted strain, did not succeed, indicating that the titre of antibodies directed solely against CREA may be too low for this approach to work. Thus the experiments outlined in this chapter allow some preliminary indications of *in vivo* binding, but definitive studies await the generation of specific antibodies.

CHAPTER 6

Chapter 6

Concluding Discussion

The aim of this study was to investigate the wide domain regulatory protein CREA of *A. nidulans* in order to obtain information on how it functions at the molecular level to bring about carbon catabolite repression. The isolation of mutant alleles with a phenotype of derepression (in the presence of glucose) suggests that the product of the *creA* gene is negatively acting. Northern analysis has shown this to be at the level of transcription for many genes under carbon catabolite control. An analysis of the theoretical protein sequence of CREA revealed the presence of two zinc fingers of the Cys_2His_2 type. This motif is present in many regulatory proteins and enables them to recognise and bind specific sequences 5' of regulated genes.

A fusion protein containing the Cys_2His_2 zinc finger region of CREA was shown in this study to bind DNA in a sequence specific manner. Binding to DNA was abolished in the presence of the zinc ion chelator, 1,10-phenanthroline, emphasising the importance of the tetrahedrally co-ordinated Zn^{2+} ion in the secondary structure of this motif. The ability of the CREA fusion protein to bind fragments from the 5' region of the *amdS, facB* and *creA* genes which are known to be under carbon catabolite control was shown in this study.

6.1 Binding studies on the 5' region of *amdS*

Within the 1.8 kb region 5' of *amdS,* strong binding by the GST-CREA fusion protein to a fragment 159 to 11 bp upstream of the start point of translation was found. DNase I sensitivity assays showed that a very large region of this fragment was protected on at least one strand, including sequences very similar to the consensus core binding site of the yeast protein, MIG1 (5' SYG GGG 3'). Within the protected region, there are at least four sub-sites similar to the MIG1 binding site consensus, and these have been called **A1**, **A2**, **A3** and **A4**. The **A1**, **A2** and **A3** sites were found to be sufficient for strong binding *in vitro*.

A fragment, PST186, located 400 to 200 bp upstream of the start point of translation of *amdS* was also found to bind the CREA fusion protein. However, when two sub-fragments from this region were tested, neither showed strong binding. The larger sub-fragment showed weak retardation, and contained a T GTG GGG sequence but no other similar sequence. The smaller sub-fragment did not contain any sequences similar to the MIG1 core binding sequence, and did not bind CREA. This result supports findings from the studies of the binding properties of oligonucleotides from the 5' region of *amdS* (section 3.8.2) that one site is not sufficient for strong binding. Why the larger fragment, PST186, bound strongly even though it contained just the one site is not known. CASTing (chapter 4) did not reveal a totally novel binding site for the CREA fusion protein and therefore perhaps strong binding to large fragments with a single site by CREA may be possible due to a non-specific, or general, attraction to DNA.

6.2 Binding studies on the 5' region of *facB*

Strong binding to fragments close to the start point of translation of *facB* was not found, but strong binding to a fragment located 1.38-1.11 kb upstream was seen. The start point of transcription of *facB* has not been determined (M.J. Hynes pers. comm.) and thus if *facB* has a long untranslated 5' region the binding to these sites up to 1.38 kb upstream may be significant. DNase I sensitivity assays show that this fragment was protected in three regions. Consistent with other results presented in this study that CREA recognises SYG GRG sequences, two of these protected regions span SYG GRG sequences. However, a third less well protected region does not. Studies using GST-ALCR have also found a region which was protected in DNase I sensitivity assays that did not correspond to the consensus binding site for the GST-ALCR protein (Kulmburg *et al.*, 1992b). Methylation interference assays showed that the guanines in this region were not protected and thus this region did not represent a specific binding site for GST-ALCR (Kulmburg *et al.*, 1992b). This region which did not contain any consensus binding sites for GST-ALCR was adjacent to a region which did. Therefore it is possible that steric interference between the DNase I enzyme and the GST fusion protein allowed protection from DNase I digestion. This does not explain why additional protection was not detected adjacent to all the observed regions of protection. The third protected region found

in this study was not GC rich, infact it was AT rich (75 %). Thus this region may span sequences able to be protected because of specific protein-DNA interactions with a putative AT binding domain present in the GST-CREA fusion protein (see below). Removing 13 bp, including the **F7** site, from the FACB1 fragment leaves one SYG GGG site (**F6**) and this fragment did not bind CREA. This is consistent with the findings that at least two binding sequences are required for strong binding *in vitro*, where one is a strong binding site and the other is either strong or weak. Two exceptions being the O2 oligonucleotide from the *prn* region and the oligonucleotide L134 both of which bind with average affinity, contain AT rich, T7 and strong core binding sequence by no other near site. There is also evidence to suggest that two sites are also required *in vivo*, since adjacent sites have been found 5' of *amdS, alcR* and within the *prn* intergenic region (this study; Kulmburg *et al.*, 1993; Cubero and Scazzocchio, 1994). There are at least two possible reasons why CREA sites are found adjacent to each other. Firstly, native CREA molecules may interact *in vivo*, and secondly, if binding to this small region spanned by CREA binding sites will prevent activation. The latter being due to inactivation of an enhancer sequence or where binding by CREA is required to prevent binding by *trans*-acting factors. If CREA fusion proteins are assumed to be interacting *in vitro*, then this leads to the suggestion that close proximity is required for strong binding. It is surprising that the fragment containing the **F6** and **F5** sites (137 bp apart) was able to bind GST-CREA weakly, and it implies an interaction by looping out of the intervening DNA. The interaction *in vitro* may not necessarily imply interactions *in vivo* since *in vitro* interactions may be due to protein-protein interactions of the GST moieties and not CREA. The theoretical CREA protein sequence does not contain any known dimerisation motifs, such as a leucine zipper. Mutants lacking the carboxy terminal, but retaining the zinc finger region, have been isolated and display a derepressed phenotype, and the phenotype of these mutants may be due to the absence of a dimerisation domain.

6.3 Binding studies on the 5' region of *creA*

Binding to the 5' region of *creA* was found and supports an autoregulatory role for the *creA* gene product (Dowzer, 1991; Dowzer and Kelly, 1991).

6.4 Analysis of the CREA binding site using CASTing

The recognition sequence for CREA was investigated further using CASTing. A core sequence of SYG GGG was present in all oligonucleotides that bound GST-CREA with average to strong affinity. Oligonucleotides which bound with weak affinity contained sites which deviate at one position from the consensus SYG GGG and thus SYG GAG sites are not as strong as SYG GGG sites. A T in position 5 and a G in position 6 was preferred two and a half times and ten times respectively over a C in either position. A preference for A's and T's 5' of this core was evident and the only oligonucleotide with a single set of 4Gs able to bind with strong affinity contained AAAAA 5' of a core T GTG GGG sequence. The lack of an AT rich region did not however prevent binding, indicating that the core sequence is sufficient for recognition by CREA but suggests that the AT rich region may allow higher affinity binding as was proposed for the MIG1 protein (Lundin *et al.*, 1994). A T in position 7 was present in 75% of selected oligonucleotides. This is very similar to the situation in *S. cerevisiae* where a T7 was present in all naturally occurring MIG1 sites 5' of the *SUC2* and *GAL* genes. MIG1 has also been shown to be involved in carbon catabolite repression of the genes of maltose utilisation including the pathway specific regulatory gene *MAL63* and the structural genes *MAL61* and *MAL62*. Binding studies on oligonucleotides from the sites present in the 5' region of these genes showed that the *MAL63* binding site bound MIG1 despite the fact that it did not contain a T7 (Hu *et al.*, 1995). When an oligonucleotide containing the naturally occurring SUC2 **A** binding site had the T7 replaced with other nucleotides strong binding still resulted. Thus there is not an absolute requirement for a T in position 7, at least *in vitro*. Perhaps a T in this position affects the kinetics of the binding reaction, something that is likely to be very important *in vivo*. This is especially the case if competition between DNA binding proteins is part of the mechanism of repression as has been suggested for CREA and MIG1.

6.4.1 Models depicting the possible configurations for the low and high mobility complexes observed

When the selected oligonucleotides were used in mobility shift analyses it was clear that some oligonucleotides formed a high mobility complex (HMC) while others formed a low mobility

complex (LMC). Upon further inspection 91% of the oligonucleotides that formed a low mobility complex contained two core binding sequences for CREA, and 90% of those forming a high mobility complex contained one strong site. The obvious explanation is that oligonucleotides with only one site are able to bind just one CREA molecule and thus migrate as a high mobility complex whereas oligonucleotides with two sites bind two molecules of CREA and thus migrate as a low mobility complex. This explanation was also put forward by workers for the *prn* intergenic region (Cubero and Scazzocchio, 1994). When the oligonucleotide L108 was tested using decreasing amounts of CREA protein, the labelled DNA was partitioned increasingly between the low mobility complex and the free probe and not between the low and high mobility complexes. Thus the observation that low mobility and high mobility complexes were not usually present simultaneously in mobility shift analyses of oligonucleotides able to form a low mobility complex, can be explained by co-operative binding, where occupation of both binding sites is more favourable than occupation of either one alone. The simplistic model, of having one CREA molecule bound in the high mobility complex and two molecules in the low mobility complex (Fig. 6.1a,b), is flawed in that it does not explain why a C in position 22 occurred at very high frequency in oligonucleotides that form a high mobility complex. Such a high frequency implies that this base is involved in specific amino acid-base pair interactions with CREA, however the only way the same CREA molecule could be able to bind at both the SYG GGG sequence and at a sequence containing a C 16 bp upstream would be if it was partitioning itself between the two sequences, where the frequency of occupancy for each site was a function of their relative affinities. This could occur, for example, through disproportionation or through sliding between the two "sites". A simpler explanation would be to have two CREA molecules bound, one at the consensus SYG GGG site and one at the sequence CGCC which involves a C in position 22. The second site would not be expected to be a strong site for CREA as it deviates significantly from the consensus. Therefore the second non consensus interaction may be stabilised by protein-protein interactions with a molecule bound to the consensus SYG GGG site (Fig. 6.1c). It is also possible that CREA binds as a dimer and/ or is able to interact with the AT rich region immediately 5' of the CGCC sequence, and that these features increase binding affinity (Fig.

Figure 6.1 Models depicting possible configurations for the high and low mobility complexes observed

Figure 6.1a

The GST-CREA fusion protein is represented by the square (= zinc finger region), ellipse (= hypothetical AT rich binding region), the GST portion of the protein (= shaded circle) and the larger circle which represents the rest of the CREA peptide. The simplistic model has one GST-CREA molecule bound to an oligonucleotide forming a high mobility complex (HMC) where binding involves interactions between the zinc finger region and the SYG GGG sequence.

Figure 6.1b

Shown are two GST-CREA molecules bound to an oligonucleotide forming a low mobility complex (LMC).

Figure 6.1c

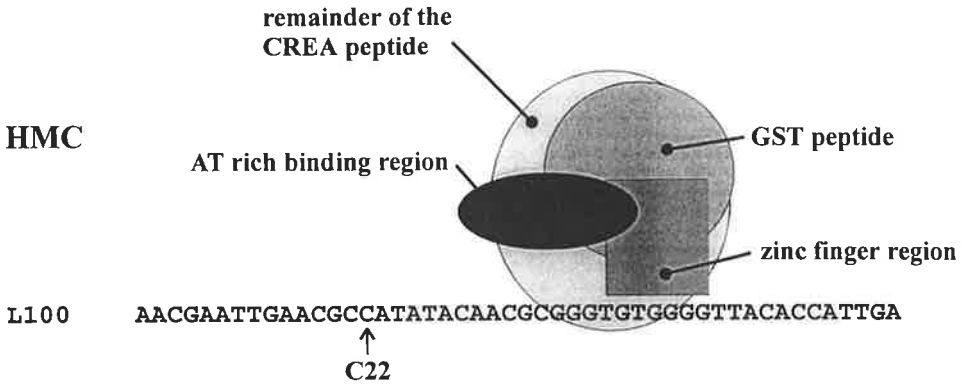
An alternative model for binding to an oligonucleotide which forms a high mobility complex. Two molecules of GST-CREA are bound. Co-operative binding is presumed to allow binding of the second molecule to the incomplete binding sequence. Interactions between the hypothetical AT rich binding region and the sequence 5' of the CGC of primer 602 may increase binding affinity and is also shown here.

Figure 6.1d

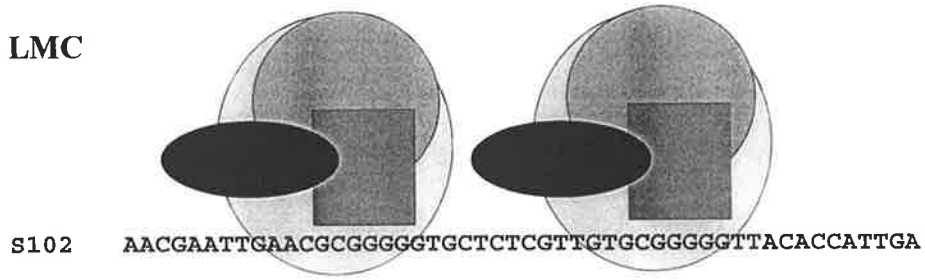
An alternative model to that of Figure 6.1b. Here two pairs of GST-CREA dimers are shown bound to a low mobility complex forming oligonucleotide.

In all cases the interacting GST-CREA proteins are illustrated as interacting via the GST domains. It is possible that these molecules interact via a region of the CREA peptide but this is not shown.

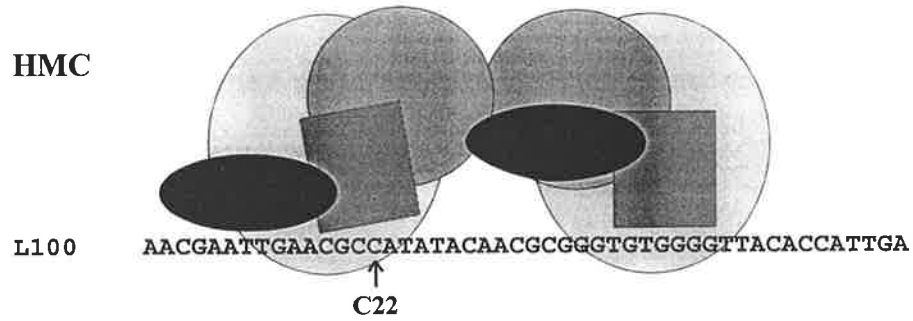
a)


HMC
remainder of the CREA peptide
GST peptide
AT rich binding region
zinc finger region
L100
AACGAATTGAACGCCATATACAACGCGGGTGTGGGGTTACACCATTGA
C22

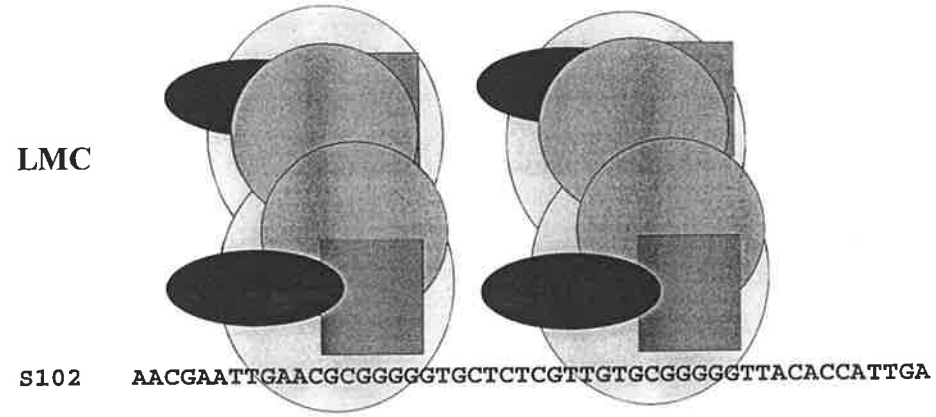
b)


LMC
S102
AACGAATTGAACGCGGGGGTGCTCTCGTTGTGCGGGGGTTACACCATTGA

c)


HMC
L100
AACGAATTGAACGCCATATACAACGCGGGTGTGGGGTTACACCATTGA
C22

d)


LMC
S102
AACGAATTGAACGCGGGGGTGCTCTCGTTGTGCGGGGGTTACACCATTGA

6.1c). The low mobility complex could result from a conformational change where the two CREA molecules now do not interact, as each has a highly favoured interaction with the DNA strand (Fig. 6.1b). Alternatively the low mobility complex could result from the binding of two dimers, one to each site, resulting in a total of four GST-CREA molecules bound to each site (Fig. 6.1d). These possibilities could be distinguished by determining the number of GST-CREA molecules present in the low and high mobility complexes. This could be achieved using a UV cross-linking experiment similar to the one which has been used to show that ADR1 binds as a monomer at low concentration but at higher concentration two molecules are bound (Thukral *et al.*, 1991). If GST-CREA was found to bind as a dimer it would be important to determine whether this interaction was via the CREA or GST domains. Alternatively full length CREA protein could be produced *in vitro*, as was done for MIG1 (Nehlin and Ronne, 1990), and used in mobility shift and UV cross-linking experiments.

6.4.2 Other sequences revealed by CASTing

CASTing revealed that in addition to the SYG GGG sequence, SYG GAG and GYG GGY sequences were present in oligonucleotides forming a low mobility complex. Work on the *prn* cluster has shown that the sequence GCG GAG (site **3.2**) is functional *in vivo* (Sophianopoulou *et al.*, 1993; Cubero and Scazzocchio, 1994). CASTing did not reveal any SYG GAG sites that were strongly binding on their own, therefore the **3.2** *prn* site is probably functional because it is located in close proximity to the SYG GGG site (**3.1**). Furthermore **3.2** contains a G6 which was found to be preferred in sites that deviate from the SYG GGG consensus (Table 4.7), and contains a T7 which was preferred for all sites (Fig. 4.3). The 5' region of the *MAL63* and the intergenic region of *MAL61* and *MAL62* genes contain multiple binding sites for MIG1 (Hu *et al.*, 1995). One of these contained a perfect AT rich region, a T in position 7 and a SYG GAG sequence yet was not able to bind MIG1 *in vitro*. This is further evidence that for CREA and MIG1 contacts made between GAG and zinc finger I are not as strong as those of GGG with zinc finger I. This can be explained by a reduced number of amino acid to base contacts, as arginine is proposed to make two hydrogen bond contacts with a guanine, one to the N7 and one

to the O6 of the guanine (Seeman *et al.*, 1976). Cubero and Scazzocchio, (1994), report that not all combinations fitting SYG GAG sequences bind CREA. They found that fragments containing some of these sequences did not bind and for others, when they were present in fragments which did bind, they were not protected. The only SYG GAG sequence which was protected was the site **3.2**. Within the fragment ALU148 of *amdS*, this study found that a large region of the DNA was protected including sequences that deviate from the consensus SYG GRG. One such sequence GAG GGT lies 18 bp downstream of the site **A1**. A similar situation occurs within the *prn* intergenic region where the pair of sites **P4** and **P5** are 15 bp downstream of site **3.1**. Cubero and Scazzocchio, (1994), report that these sites were not protected in DNase I sensitivity assays and methylation interference assays, and that therefore binding by CREA is very context dependent. From CASTing it was found that one oligonucleotide, S119, did not contain two obvious CREA binding sites but yet migrated as a LMC. Conversely additional potential weak CREA binding sites can be found in oligonucleotides which form a HMC. Therefore the interaction of CREA with DNA appears to be very context dependent.

Work on MIG1 from *S. cerevisiae* showed that an oligonucleotide in which G1 was changed to a T (GYG GGT) binds MIG1, but does so extremely poorly (Lundin *et al.*, 1994). It would be interesting to determine whether MIG1 is capable of binding oligonucleotides with two binding sites with a higher affinity than it binds oligonucleotides with just one site. It would also be of interest to determine whether or not strong binding resulted even if the second site deviated from the MIG1 consensus. If MIG1 does not tolerate sites which deviate from its consensus sequence then this may represent a difference between the two proteins in the way that they bind DNA, or may simply mean that CREA tolerates mismatches much more readily than does MIG1. The binding sites for MIG1 have all been fairly well conserved whereas those found 5' of the genes regulated by CREA appear to diverge from the preference for AT rich, T7 and SYG GGG sequences.

An alignment of *in vivo* binding sites shows that a C in position 22 does not occur more frequently than at random. It is therefore proposed here that it was not a "C22" which was important for binding to oligonucleotides (chapter 4) but rather the presence of a sequence which allowed some binding by another molecule of GST-CREA which gave these oligonucleotides higher affinity for the GST-CREA fusion protein. The affinities of the sites determined using the *in vitro* generated GST-CREA fusion protein may not reflect the affinities of the native CREA protein. The most popular triplet 5' of the 4G string was determined using CASTing to be TGT, but the majority of sites located 5' of the genes studied so far did not contain this triplet (Figures 3.21-3.26).

6.5 The presence of sequences fitting the CREA consensus binding site within the 5' region of genes present in the data base

Knowing that CREA recognises SYG GGG, SYG GAG and possibly GYG GGY sequences (albeit with varying affinities) it is possible to search the data base of *A. nidulans* genes for these sequences. Using the gcg "findpatterns" computing program[1] (Smithies *et al.*, 1981), the SYG GRG and GYG GGY sites present in the 5' region of cloned *A. nidulans* genes were identified (Fig. 6.2a-d). Strong sites are defined as those containing a string of 4G's (SYG GGG) and weak sites those fitting the sequence SYG GAG. Although the significance of sequences fitting the sequence GYG GGY is not known, these sites were also included in the analysis. The orientation of a sequence is denoted by a rectangle for sites where the G rich strand is the coding strand and an ellipse for when the G rich strand is the non-coding strand. An analysis with respect to the type of site (strong vs weak), its orientation, the total number of sites and whether any are adjacent, was carried out.

6.5.1 Genes which are expected to be regulated by CREA

There are a number of genes (Fig. 6.2a) which are involved in the utilisation of carbon sources that are alternatives to glucose and thus are expected to be repressed by CREA, indeed many

[1]see also the Program Manual for the Wisconsin Package, Version 8, September 1995, GCG, 575 Science Drive, Madison Wisconsin, USA, 53711.

of them have been shown to be regulated by CREA at the level of transcription (section 1.2.3). These genes in general contain many strong theoretical sites (Fig. 6.2a). In random DNA sequence the expected frequency for a CREA site of the types fitting the consensus SYG GRG is one in every 512 bp. Therefore all of these genes contain a higher than expected frequency of theoretical CREA sites. Except for *aldA*, for which only a small amount of sequence 5' is available, these genes all contain at least two strong sites. In addition, except *facA* and *aldA*, they contain a pair of theoretical sites within 50 bp of each other where at least one of these is a strong theoretical site. Failure to find such a pair of theoretical sites 5' of *facA* may be due to the relatively small amount of sequence 5' of the start point of transcription of this gene available in the data base. These results are consistent with the importance of a pair of sites in the regulation of these genes by CREA. Only one of these genes, *alcR*, contains a strong theoretical site between the start point of transcription and the start point of translation and therefore repression of transcription by blocking elongation by RNA polymerase does not appear to be a common mechanism. The regions 5' of these genes contain sites in both orientations and there does not appear to be a rule for what sort of sites and what orientation of these sites is acceptable. The region 5' of *acuD* indicated in Figure 6.2a was able to confer carbon catabolite repression on a heterologous gene in a *creA* dependent manner, and it contains two SYG GRG sequences (De Lucas *et al.*, 1994b). This suggests that this pair of sites, which lie between -356 and -375, are the likely *in vivo* functional sites for CREA 5' of *acuD*. The presence of pairs of theoretical sites 5' of these genes and also the fact that pairs of sites have been implicated in the *in vivo* analysis of *amdS, alcR* and *prn* suggests that a common mechanism of repression involves at least two sites.

6.5.2 Genes which are not expected to be regulated by CREA

Some genes are not expected to be regulated by CREA but yet an analysis of the 5' region of these genes shows that they do contain potential CREA binding sites (Fig. 6.2b,c).

Most of the genes shown in Figure 6.2b are involved in carbon flux, but are not expected to be regulated by CREA. Also included in Figure 6.2b are the *argB* and *pyrG* biosynthetic genes.

Figure 6.2a-d Position and type of sequences fitting the CREA binding site, present in the 5' region of *A. nidulans* genes

The gcg "findpatterns" program was used to locate sequences fitting the recognition site for CREA within the 5' region of *A. nidulans* genes present in the database (as of December, 1995 and including *alcB*). The 5' regions of various genes are shown by the horizontal lines relative to the start point of translation (+1). Where known the start point of transcription is shown by an arrow(s). Sites where the G rich strand is the coding strand are shown as rectangles whereas sites where the G rich strand is the non-coding strand are shown as ellipses. The position of strong CREA binding sites fitting the sequence SYG GGG is shown by black shading. The position of weaker sites fitting the sequence SYG GAG (light shading) and those fitting the sequence SYG GGY (unshaded) are also shown. The fragment from the 5' region of *acuD* which was able to confer carbon catabolite repression to a heterologous gene is shown with a dotted line. For the *facB, facA, acuD, acuE* and *amdS* genes the location of FACB binding sites and putative binding sites are shown by letters indicating strong binding (S), medium binding (M), weak binding (W) or putative site (ND). A description of the gene product, the accession number and references for each gene is given below.

creA encodes a regulatory gene involved in carbon catabolite repression, accession number L03563 (Dowzer and Kelly, 1991; Shroff *et al.*, submitted).

facB encodes a regulatory gene involved in acetate induction, (Katz and Hynes, 1989b; Todd *et al.*, submitted). Start point of transcription is not known (M.J. Hynes pers. comm).

facA: structural gene for acetyl CoA synthase, accession number X16990 (Connerton *et al.*, 1990).

acuD encodes isocitrate lyase, accession number X62696 (Gainey *et al.*, 1992; Bowyer *et al.*, 1994).

acuE encodes malate synthase, accession number X56671 (Sandeman *et al.*, 1991).

amdS encodes an acetamidase, accession number M16371 (Corrick *et al.*, 1987)

alcR encodes a regulatory gene involved in induction of *alcA* and *aldA* by ethanol, accession number M24071 (Felenbok *et al.*, 1988; Kulmburg *et al.*, 1991).

alcA encodes alcohol dehydrogenase I, accession number M16196 (Gwynne *et al.*, 1987)

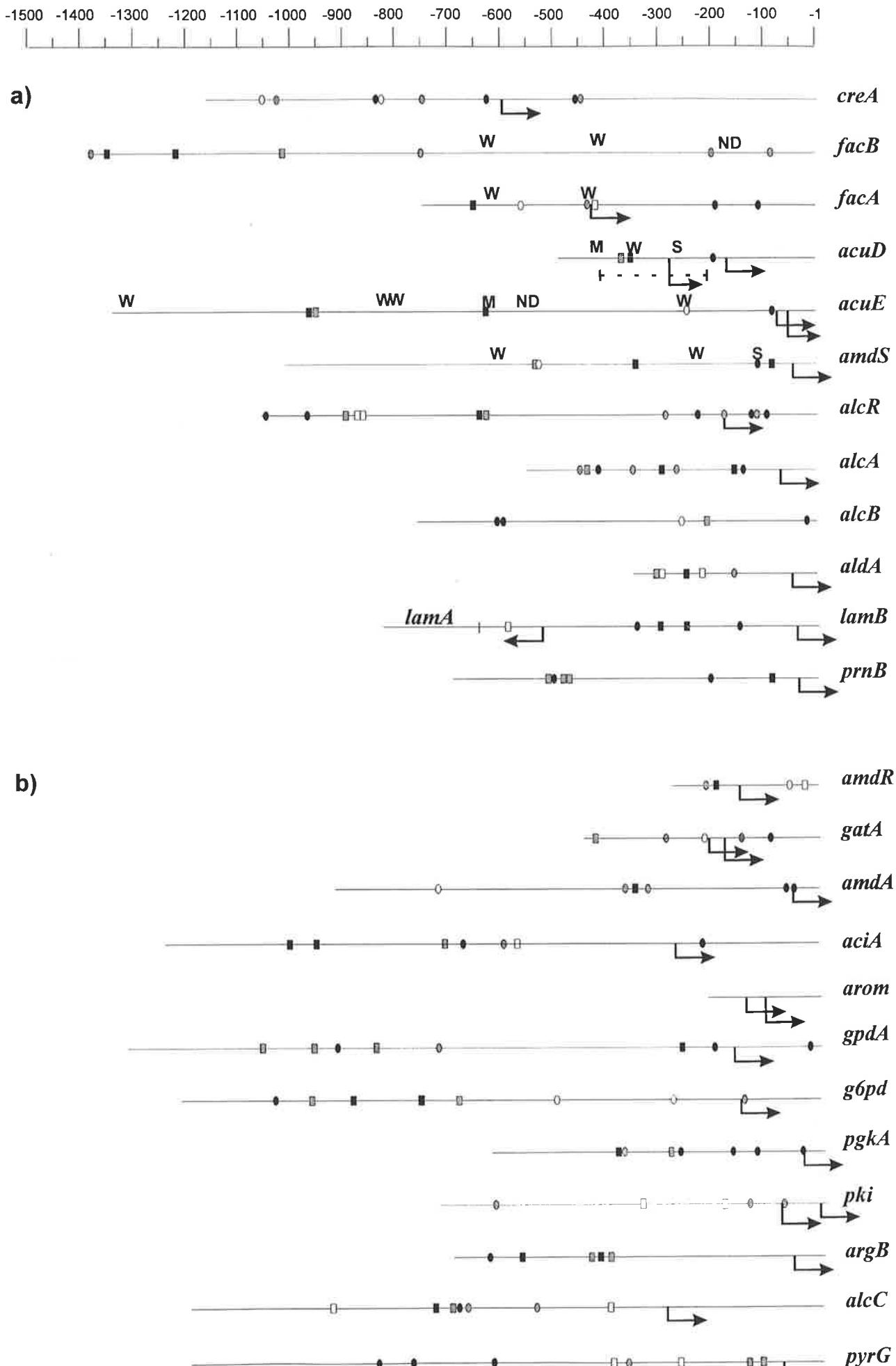
alcB encodes alcohol dehydrogenase II (Hunter *et al.*, 1996).

aldA encodes aldehyde dehydrogenase, accession number M16197 (Pickett *et al.*, 1987; Gwynne *et al.*, 1987).

lamA/lamB encode genes involved in the metabolism of lactams, accession number M77283 (Richardson *et al.*, 1992).

amdR encodes a regulatory gene involved in omega amino acid induction of the *gatA* and *lamA*, *lamB* genes and also regulation of *amdS*, accession number M31517, M36111 (Andrianopoulos and Hynes, 1990).

gatA encodes GABA transaminase, accession number X15647 (Richardson *et al.*, 1989).


-1500 -1400 -1300 -1200 -1100 -1000 -900 -800 -700 -600 -500 -400 -300 -200 -100 -1
a)
creA
facB
W W ND
facA
W W
acuD
M W S
acuE
W WW M ND W
amdS
W W S
alcR
alcA
alcB
aldA
lamA
lamB
prnB
b)
amdR
gatA
amdA
aciA
arom
gpdA
g6pd
pgkA
pki
argB
alcC
pyrG

amdA encodes a regulatory protein involved in acetate regulation of *amdS* and *aciA*, accession number L28810 (Lints *et al.*, 1995)

aciA encodes a putative formate dehydrogenase, accession number Z11612, S51247 (Saleeba *et al.*, 1992).

***arom* locus** encodes the penta functional polypeptide catalysing steps 2-6 in the biosynthesis of prephenate, a precursor for aromatic amino acids, accession number X05204, X03180, X03684 (Hawkins, 1987; Charles *et al.*, 1985, 1986).

gpdA encodes glyceraldehyde-3-phosphate dehydrogenase, accession number M19694, M22758 (Punt *et al.*, 1988)

g6pd encodes glucose-6-phosphate dehydrogenase, accession number X84001 (Schaap, P.J., Muller, Y. and Visser, J., unpublished).

pgk encodes 3-phosphoglycerate kinase, accession number M27549 (Clements and Roberts, 1986).

pki encodes pyruvate kinase, accession number M36918 (de Graaff and Visser, 1988).

argB encodes ornithine carbamoyltransferase, accession number I00291, M29819 (Upshall *et al.*, 1986).

alcC encodes alcohol dehydrogenase III, accession number X02764 (McKnight *et al.*, 1985).

pyrG encodes orotidine-5'-phosphate decarboxylase, accession number M19132 (Oakley *et al.*, 1987)

areA encodes a regulatory gene involved in nitrogen metabolite repression, accession number X52491(Kudla *et al.*, 1990; Langdon *et al.*, 1995, 1996)

nirA encodes a regulatory gene involved in nitrate assimilation, accession number M68900 (Burger *et al.*, 1991).

crnA encodes nitrate permease, accession number M61125 (Unkles *et al.*, 1991a,b).

niiA encodes nitrite reductase, accession number M58289 (Johnstone *et al.*, 1990).

niaD encodes nitrate reductase, accession number M58291 (Johnstone *et al.*, 1990).

uaY encodes a regulatory gene involved in purine utilisation, accession number X84015 (Suarez *et al.*, 1995)

uapA encodes uric acid-xanthine permease, accession number X71807 (Gorfinkiel *et al.*, 1993)

uapC structural gene encoding a purine permease, accession number X79796 (Diallinas *et al.*, 1995)

uaZ encodes urate oxidase, accession number X72210 (Oestreicher and Scazzocchio, 1993)

hxA structural gene for xanthine dehydrogenase, accession number X82827 (Glatigny and Scazzocchio, 1995)

fluG encodes a gene required for production of an extracellular developmental signal, accession number L27817 (Lee and Adams, 1994).

brlA the alpha and beta transcripts encode regulatory proteins involved in conidiophore development, accession number M20631, L25858, L24076 (Adams *et al.*, 1988; Prade and Timberlake, 1993; Adams and Han, unpublished 5' sequence).

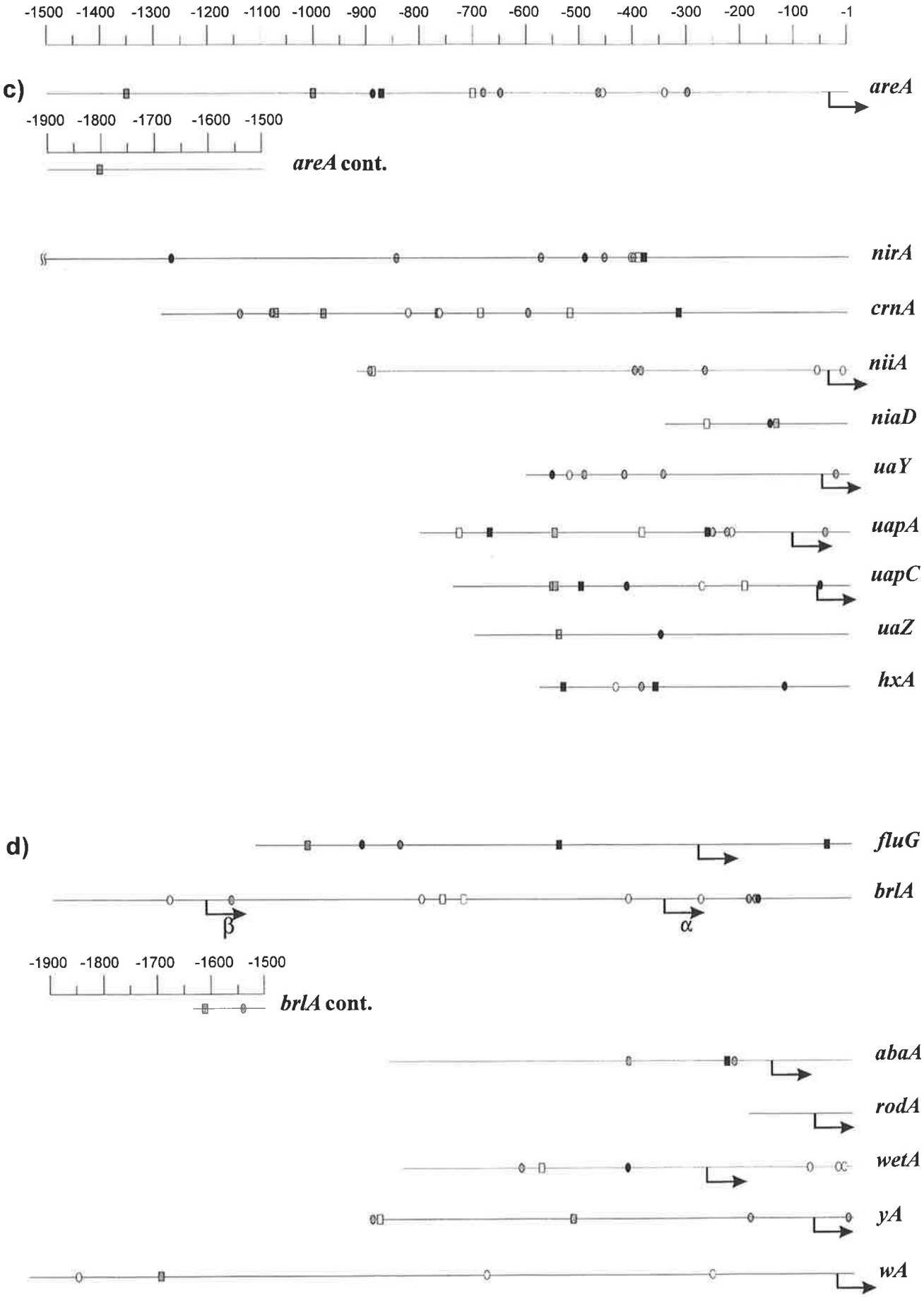
abaA encodes a regulatory gene involved in conidiophore development, accession number J04850 (Mirabito *et al.*, 1989).

rodA encodes a rodlet peptide involved in conidiophore development, accession number M61113 (Stringer *et al.*, 1991).

wetA encodes a regulatory gene involved in formation of the conidiophore, accession number M60528, M35758 (Marshall and Timberlake, 1991).

yA encodes laccase, accession number X52552 (Aramayo and Timberlake, 1990).

wA encodes a putative polyketide or fatty acid synthase, accession number X65866, S51229 (Mayorga and Timberlake, 1992).


-1500 -1400 -1300 -1200 -1100 -1000 -900 -800 -700 -600 -500 -400 -300 -200 -100 -1
c)
areA
-1900 -1800 -1700 -1600 -1500
areA cont.
nirA
crnA
niiA
niaD
uaY
uapA
uapC
uaZ
hxA
d)
fluG
brlA
β
α
-1900 -1800 -1700 -1600 -1500
brlA cont.
abaA
rodA
wetA
yA
wA

There are only six fewer strong binding sites per base pair of sequence present in the 5' sequences of these genes than there are within the 5' regions of the genes shown in Figure 6.2a. The *amdR* gene which codes for a positively acting regulatory protein involved in the induction of *amdS* by ω-amino acids has within its 5' sequence a pair of theoretical CREA binding sites. Northern analysis has shown that this gene is constitutively expressed with respect to the presence of the inducer (Andrianopoulos and Hynes, 1988). However no studies on *amdR* mRNA levels in the presence of repressing and derepressing carbon sources or the effect of various *creA* mutations on mRNA levels has been reported. The *amdR* gene from *A. oryzae* has been cloned and sequenced and it also contains theoretical binding sites for CREA, three of which are within 50 bp of each other (Wang *et al.*, 1992). Therefore it can not be ruled out that CREA does have some effect on the expression levels of the *amdR* gene although it seems unlikely that this gene is subject to strong carbon catabolite repression. Another gene, *argB*, (Fig. 6.2b) encodes the enzyme required to convert ornithine to citrulline in the biosynthetic pathway of arginine (Upshall *et al.*, 1986). It is not known to be regulated by carbon source yet it contains three strong theoretical binding sites and at least one pair of theoretical binding sites. Other genes encoding glycolytic enzymes such as *pgkA,* (encoding 3-phosphoglycerate kinase) and *gpdA* (encoding glyceraldehyde 3-phosphate dehydrogenase) are not expected to be repressed by CREA yet they contain many strong theoretical CREA binding sites. The regulation of expression of these and other genes involved in carbon metabolism is reviewed by Hondemann and Visser, (1994). Although MIG1 is best known for its role as a repressor, it was found to activate transcription in an *ssn6* background (section 1.2.2.3d, Treitel and Carlson, 1995). For this reason it is possible that CREA is able to act as both a repressor and an activator, and thus a positive role for CREA in the regulation of these genes cannot be ruled out. Regulation of *pgkA* and *gpdA* by a gluconeogenic carbon source results in an increase in expression, however, the sequence responsible for this, the *gpd* box, is not related to the CREA binding site and is not involved in carbon catabolite repression (Streatfield *et al.*, 1992; Punt *et al.*, 1990, Punt *et al.*, 1992). The pyruvate kinase gene (*pki*) in comparison does not have any strong theoretical sites within the 5' sequence available thus far yet it is also a major point of

control for glycolysis. The glucose-6-phosphate dehydrogenase gene (*g6pd*) is not expected to be repressed by CREA. It is involved in directing the flux between the pentose phosphate pathway and glycolysis and is more dependent on the type of nitrogen source available (reviewed by Hondemann and Visser, 1994). As its 5' sequence does not contain a pair of theoretical binding sites, this is in keeping with the expected lack of direct regulation of this gene by CREA. Expression of the *alcC* gene is not repressed by glucose, yet it too has a number of theoretical CREA binding sites (Fig. 6.2b; Kelly *et al.*, 1990). Northern analysis has shown that the *alcC* message increases when grown under conditions of carbon catabolite repression, thus it can not be ruled out that CREA has a positive effect on the expression of this gene.

The genes involved in nitrogen utilisation (Fig. 6.2c) are not expected to be regulated by CREA. These genes have less theoretical CREA binding sites than the genes expected to be repressed by CREA and of those present the weaker affinity sites predominate (Fig. 6.2c). However, at least five of these genes contain a pair of theoretical CREA binding sites, *areA*, *nirA*, *niaD*, *uapA* and *hxA*. There is no reason to expect that these genes are regulated by CREA. However, as these genes have never been directly tested, a negative or a positive role for CREA in the regulation of these genes has not formally been ruled out. There must be a mechanism to prevent CREA from binding the sites 5' of genes not regulated by CREA, or a mechanism that prevents it repressing gene expression if it does bind.

6.5.3 Genes involved in conidiophore development

Strains containing *creA* mutations take longer to conidiate on complete and also to a lesser extent on minimal media. Therefore CREA may be involved in the regulation of conidiophore development. The signals which trigger conidiation in *A. nidulans* are poorly understood but include reduced nutrient availability (Saxena and Sinha, 1973, cited by Timberlake, 1993), red light (Mooney and Yager, 1990) and exposure to an air-surface interface (Axelrod, 1972). The *fluG* gene is involved in the synthesis of a low molecular weight, diffusible signal required to trigger conidiophore development by an air-surface interface (Lee and Adams, 1994). This gene

triggers development independently of the nutritional status and therefore is probably not affected by CREA. The 5' sequence of this gene contains three strong theoretical CREA binding sites but none of these form a pair of sites (Fig. 6.2d). This is consistent with the expected lack of involvement of the *creA* gene product in the regulation of this gene. The general lack of theoretical CREA binding sites present in the genes expressed late in conidiophore development, *rodA*, *wetA*, *yA* and *wA* correlates well with the fact that these genes are not anticipated to be regulated by CREA. It is possible that CREA directly represses transcription of the *brlA* and *abaA* genes since a pair of theoretical binding sites are present in the 5' region of these genes (Fig. 6.2d). As these genes are involved in the early stages of conidiophore development this may reflect a true role for the CREA protein in their regulation. A search for genes which act earlier in conidiophore development to the *brlA* gene is underway (Marhoul and Adams, 1995) and this may provide more information on whether CREA is involved either directly or indirectly in the regulation of conidiophore development.

6.5.4 Genes from *A. niger*

The CREA proteins from *A. niger* and *A. nidulans* are 90 % similar at the amino acid level and over the zinc finger region the two proteins are 96% identical including the positions -1, +3 and +6 (Drysdale *et al.*, 1993). For this reason these proteins are predicted to recognise identical binding sites. The position and type of SYG GRG sequences 5' of a number of genes which are present in the data base has been determined. Like the situation found for *A. nidulans*, potential CREA binding sites in the form of a pair of sites where at least one is a strong site (SYG GGG) were found. Two genes, *aldA* and *suc1*, which could be expected to be regulated by CREA contain at least one pair of potential CREA binding sites consistent with a role for CREA in the regulation of these genes (O'Connell and Kelly, 1989; Boddy *et al.*, 1993). A pair of theoretical sites was also found within the *niiA-niaD* intergenic region, and these genes are not expected to be regulated by CREA (Unkles *et al.*, 1992). The small amount of sequence 5' of the *pyrG* gene of *A. niger* contains two pairs of theoretical sites even though this enzyme is not expected to be regulated by CREA (Wilson *et al.*, 1988). Interestingly these sites are not present in the *A. nidulans* homologue (Fig. 6.2b), indicating that they are not functionally conserved.

6.5.5 Conclusions

All *in vivo* analyses confirm the importance of a pair of sites where one site is a strongly binding site. Although the absence of sites 5' of a given gene is almost certainly an indication that CREA does not directly regulate the gene, the reverse may not be true as there are pairs of sites present within the 5' region of genes not expected to be regulated by CREA. As these genes have not been tested for an effect of *creA* on their expression a direct role for CREA can not be ruled out. As mentioned in section 3.9.5.3, multiple GST-CREA binding sites 5' of the *ipnA* gene have been found, yet deletion of the CREA site within a region responsible for carbon regulation had no effect on expression of a 5' *ipnA-lacZ* fusion construct. This gene is not thought to be repressed either directly or indirectly by CREA (Espeso and Penalva, 1994; Espeso *et al.*, 1993). Thus the presence of strong theoretical sites does not necessarily indicate a functional role for these sites *in vivo* as is probably the case for the theoretical sites found 5' of genes not expected to be regulated by CREA. The presence of sites 5' of some *A. niger* genes is consistent with a role for CREA in their regulation but like the situation for *A. nidulans* genes, potential binding sites were found 5' of genes not expected to be regulated by CREA.

The presence of paired binding sites for regulatory genes, either as direct or inverted repeats has been found for many DNA binding proteins including AREA and ABAA (Peters and Caddick, 1994; Andrianopoulos and Timberlake, 1994). Of special interest is the Cys_2His_2 containing zinc finger protein, BRLA (see also Fig. 1.3), which recognises sites 5' of the structural gene *yA* (Aramayo and Timberlake, 1993). This DNA binding protein shares the same amino acid residues as CREA at the three critical positions (-1, +3 and +6) for zinc finger I but differs at two of these positions for zinc finger II. The consensus binding site for BRLA is MRA GGG (Chang and Timberlake, 1993) which confirms the model of DNA-protein interactions shown in Figure 3.18a. BRLA appears to bind to more than one site 5' of *yA* and these are of both the inverted repeat and the direct repeat type of sequences (Aramayo and Timberlake, 1993).

In addition to the sites found in this analysis there may be sites that deviate from the consensus sequences at only one position (as was found between the two *amdS* sites **A1** and **A4**). Such sites are likely to be functional when adjacent to strong binding sites.

6.6 How does CREA function at the molecular level?

There are a number of ways that a protein could act as a repressor to bring about a decrease in the frequency of transcription initiation. For example:

1) by binding to sites that are in close proximity to or overlapping with those recognised by transcriptional machinery thereby preventing transcription initiation.

2) by binding to the promoter region to directly repress transcription initiation by interacting with the transcriptional machinery regardless of the presence or absence of an activator.

3) by competing with an activator for binding to the promoter region resulting in the inability of the activator to bind.

4) when an activator and the repressor are bound the repressor may prevent the activator from making the right contacts necessary for activation either through steric interference or via negative interactions with the activator.

5) by interaction with the activator prior to binding to the DNA resulting in the inability of the activator to bind and activate transcription.

6) by binding to an enhancer element required for high levels of transcription.

7) by recruiting a general repressor protein complex to carry out the repression.

8) by associating as part of a large complex which carries out many functions, including DNA binding and repression.

The location of CREA binding sites 5' of genes that are expected to be repressed directly by CREA do not generally overlap with the start points of transcription and indeed are sometimes located a considerable distance away. Therefore, the first mechanism of repression is not likely to be a major feature of repression by CREA.

A major question that is unresolved is how the carbon status of the cell is signalled to allow CREA to effect repression. The mechanism could be due to a direct interaction of glucose (or a metabolite or small effector molecule) with CREA. This may affect the sub-cellular location of CREA or be required for binding to DNA as a dimer or multimer (this process could require DNA). Alternatively, it may cause a conformational change such that a repression domain which was concealed is now exposed, or perhaps the conformational change increases the potency of the repressor domain. The small effector molecule may be present (albeit at a low level) during growth on derepressing carbon sources and allow some repression by CREA even under these conditions. This is the simplest explanation for the results of Fillinger *et al.*, (1995) who found that under growth conditions generally considered derepressing message levels of the *alcA, alcR* and *aldA* genes were affected by the presence of the *creA30* allele. Alternatively, a more complex signalling pathway may exist if the situation for CREA is similar to that of MIG1 in yeast. Glucose is likely to cause a regulatory cascade where CREA recruits other proteins which are modified depending on the presence of glucose. There are many possible mechanisms for this type of indirect regulation including control by proteolysis (as is suggested for pH regulation in *A. nidulans*, Orejas *et al.*, 1995), phosphorylation/ dephosphorylation (as is suggested for many *S. cerevisiae* regulatory proteins such as MIG1, CAT8, GAL4; Treitel and Carlson, 1995; Hedges *et al.*, 1995; Parthun and Jaehning, 1992) and protein-protein interactions. Given the situation for carbon catabolite repression in *S. cerevisiae* (section 1.2), control by phosphorylation could be likely, although no equivalent for the *SNF1* gene has thus far been isolated from *A. nidulans*.

6.6.1 Analysis of *creA* mutants

The structure of regulatory proteins such as CREA are modular, containing discrete domains responsible for an aspect(s) of function. For this reason an analysis of mutations isolated at the *creA* locus will provide information on the domains present within CREA. Cloning and sequencing of many of the alleles containing mutations in the *creA* gene has been carried out (Table 5.1; Shroff *et al.*, submitted; R.A. Shroff and J.M. Kelly pers. comm.). The molecular

changes of mutants with a phenotype of derepression include point mutations within the zinc finger region proving that DNA binding is required for repression by CREA. An indirect method of repression such as 5) above is thus not likely to be a key component of the mechanism of repression for CREA. Two alleles resulting in temperature sensitivity, *creA204* and *creA225,* are both due to non-conservative changes of an aspartic acid at position +2 of the second zinc finger (Shroff *et al.*, submitted). The aspartic acid is well conserved amongst Cys_2His_2 zinc finger proteins (Fig. 1.3) and was shown to hydrogen bond with the amino acid at the +3 position in crystallographic studies of Zif268, providing stability to the zinc finger-DNA interaction (Pavletich and Pabo, 1991). For these reasons the temperature sensitivity of these two alleles is probably due to a reduced stability in the second finger which reduces the efficiency of the protein-DNA contacts. The *creA1* allele is due to a mutation within the linker region and does not result in a temperature sensitive phenotype. A mutation which is likely to change the sequence specificity of the recognition site has thus far not been found. A GST-CREA fusion protein which contains two amino acid changes within the zinc finger region (non conservative amino acid changes at +1 and +2 of the second zinc finger) and one outside the zinc finger region was expressed in *E. coli* and used in mobility shift assays. Binding to the ALU148 fragment from *amdS,* and to several oligonucleotides which bind the wild type GST-CREA fusion protein, was abolished by these amino acid changes (results not shown). This confirms the importance of amino acids in positions that do not directly contact DNA bases in binding affinity.

In addition to mutations within the zinc finger region, many frame shift mutations resulting in carboxy terminal truncations and alterations have been found to cause derepression. These mutations affect the acidic and the RGR1 similar domains of CREA and strains containing these mutations do not display temperature sensitive phenotypes (Shroff *et al.*, submitted). CREA proteins encoded by strains carrying these mutations may be less stable, unable to dimerise, may not respond to the signal produced in response to glucose availability, may lack a repression domain or lack a domain which is required to recruit a repressor protein(s). They

may all contain nuclear localisation domains since a candidate for the nuclear localisation domain is located immediately carboxy to the zinc finger DNA binding domain (Ostling *et al.*, 1996; see below). An example of this type of mutant is the *creA30* containing strain. This strain produces a truncated CREA peptide which could be expected to be localised in the nucleus and able to bind DNA. This is consistent with the binding seen by nuclear extracts from a *creA30* containing strain, although antibodies able to detect CREA *in vivo* are required to prove that the binding seen was due to CREA. A *creA30* strain is derepressed for alcohol dehydrogenase as measured by its inability to grow in the presence of allyl alcohol and glucose. Together with the fact that all mutations isolated thus far localised outside of the zinc finger region also cause derepression of alcohol dehydrogenase, this suggests that straightforward competition for binding sites is not the only mechanism of repression acting on the *alcA* and *alcR* genes (see below).

6.6.2 Functional analysis of the MIG1 protein

A functional dissection of the MIG1 protein from *S. cerevisiae* has been carried out and it has shown that MIG1 contains an effector domain, two domains which regulate MIG1 activity and a putative nuclear localisation domain (Ostling *et al.*, 1996). The carboxy terminal 24 amino acids which form the effector domain is active even when only fused to the zinc finger region, albeit not quite as effectively as the wild type protein. The *MIG1* gene from *Kluveromyces lactis* has been cloned and sequenced and was found to have little overall sequence identity and only 55% overall similarity to the *S. cerevisiae MIG1* gene (Cassart *et al.*, 1995). It contains two highly conserved zinc fingers and heterologous complementation has shown that the *K. lactis MIG1* gene was able to function in *S. cerevisiae* (Cassart *et al.*, 1995). Sequence comparisons between MIG1 from *K. lactis* and *S. cerevisiae* shows that a region of the effector domain is conserved between the two species. This region contains leucine-proline repeats and shares similarity to the SH3 binding domain found in many intracellular signalling proteins (Yu *et al.*, 1994 and references therein) but differs significantly from it (Ostling *et al.*, 1996). MIG1 recruits the repressor complex SSN6/ TUP1 to carry out repression and therefore this effector

domain is proposed to interact with this complex. Sequence comparisons between the effector domain of MIG1 from *K. lactis* and *S. cerevisiae* with CREA from *A. nidulans* shows that similar leucine-proline repeats occur within two separate regions of CREA (see Figure 6.3a; Ostling *et al.*, 1996). Given that MIG1 from *S. cerevisiae* and CREA have little sequence similarity except over the zinc finger domain, these stretches of similarity seem quite significant. The region from 371 to 385 contains the highest sequence similarity and is located within the carboxy terminal 50 amino acids of CREA. This is very similar to the location of the effector domains of MIG1 from both species. This region is highly conserved in the species *A. niger* (Fig. 6.3a) and would be missing in many truncation mutants, such as *creA218,* which show derepression (Shroff *et al.*, submitted). These mutants are not as extreme as some of the zinc finger mutants. It is possible that there is more than one effector domain, each required for interactions with different protein complexes and that therefore the other stretch of similarity forms perhaps another effector domain. The leucine proline repeat structure of the 239-253 stretch of sequence similarity is present in the *A. niger* sequence although the spacing does not appear to be very well conserved (Fig. 6.3a). Results which support the notion that CREA represses different loci by a different mechanism, and perhaps recruits different protein complexes to do the job, comes from the analysis of a strain containing a deletion from amino acid 361 onwards (Fig. 6.3b). This strain shows derepression for the *prn* genes but not the *alc* genes (Scazzocchio *et al.*, 1995).

Within MIG1 two regulatory domains were found, R1 and R2. R1 contains a motif known to be a core target for serine-specific protein kinases. Construction of various fusions of MIG1 with the activation domain from viral protein VP16 showed activation of a synthetic promoter containing MIG1 binding sites. The effect of multiple binding sites was synergistic, a single site producing 5 fold activation, two sites producing 20 fold activation and three sites producing 40 fold activation. Using this reporter system SNF1 dependent inhibition of MIG1 binding in the absence of glucose was shown. This inhibition was dependent on the presence of the R1 and to a lesser extent the R2 domain. Deletion of the R1 domain results in a MIG1 protein which is still able to bind DNA even in the absence of glucose thus causing a constitutively active

MIG1-VP16 protein (Ostling *et al.*, 1996). This suggests that SNF1 plays a direct role in the inactivation of MIG1 (compared to the indirect role depicted in Fig. 1.2). Removal of another domain was found to abolish activation by the MIG1-VP16 protein and this domain was called the basic domain. A search for similarity to proteins within the data base found that this domain had sequence similarity to the nuclear localisation signal of the *S. cerevisiae* protein SWI5 (Moll *et al.*, 1991; Ostling *et al.*, 1996). It has high sequence similarity with a region from CREA and directly follows the zinc finger motif in both MIG1 and CREA proteins (Fig. 6.3a). CREA is expected to be localised in the nucleus and given the conservation of the position within the protein sequences for this motif, this could be a functional domain in CREA also. A functional dissection of the CREA protein is required to confirm or deny the existence of the domains suggested by these results and such a study is currently underway in our laboratory. The presence of a nuclear localisation motif argues against control of CREA function being at the level of entry into the nucleus, although a mechanism whereby the functionality of this domain is altered by different growth conditions is possible. MIG1 appears to be subject to control by phosphorylation (within the R1 and/ or R2 regions) probably directly mediated by the SNF1 protein kinase (Ostling *et al.*, 1996). A similar form of regulation may occur for CREA. No genes have been implicated in a similar role to SNF1 from *S. cerevisiae* in *A. nidulans*, however protein kinases from higher eukaryotes have been found to have sequence similarity to SNF1 from yeast and thus this form of regulation has been conserved (see also section 1.2.2.2a).

An activation domain for the MIG1 protein was not discovered in the functional analysis carried out by Ostling *et al.*, (1996). This is because the strong activation by MIG1 reported by Treitel and Carlson (1995) was seen in an *ssn6* background. Ostling *et al.*, (1996) did not extend their analysis to such a background and therefore it is not yet known where this domain is located.

6.6.3 Comparisons on the regulation of genes regulated by CREA

Many systems regulated by CREA have been studied extensively including the alcohol regulon, proline utilisation cluster and some genes involved in the utilisation of compounds which are metabolised via acetate.

Figure 6.3a Sequence similarity of CREA with the putative effector and nuclear localisation domains of MIG1

An alignment of the putative effector and nuclear localisation domains of MIG1 from *S. cerevisiae* and *K. Lactis* with potentially similar regions of CREA from *A. nidulans* and *A. niger* (Drysdale *et al.*, 1993) and with ROX1 from *S. cerevisiae* is shown. This is a modification of Figure 2 from Ostling *et al.*, (1996). The two regions from *A. nidulans* are distinguished by the amino acid positions in the brackets. Amino acid similarities as shown in Ostling *et al.*, (1996) are shown in bold type.

Figure 6.3b Position of putative domains within the amino acid sequence of CREA from *A. nidulans*

The amino acid sequence of CREA from *A. nidulans* is shown. The zinc finger regions are boxed and the linker region between the zinc fingers is underlined. The putative nuclear localisation domain is shown with a dotted line, and the putative effector domains are indicated with dashed lines. The end point for the CREA deletion discussed in the text is marked with an arrow (Scazzocchio *et al.*, 1995).

a)

Alignments with the effector domain of MIG1

MIG1 *S. cerevisiae*	L P P I R S L P L P F P H M D
MIG1 *K. Lactis*	L P S L R S L D L L P P K
CREA *A. nidulans* (371-385)	E G A Q R K L P I P Q V P K V
CREA *A. niger*	D G R Q R K L P V P Q V P K V
ROX1 *S. cerevisiae*	Y M V S R S L S G L P L T H D
ROX1 *S. cerevisiae*	D K T A R D L P Q L S S Q L N
CREA *A. nidulans* (239-253)	R H G S R G L P S L S A Y A I
CREA *A. niger*	R H G S R - L P L L A A Y A I
	H G

Alignment of the putative nuclear localisation domain of MIG1

MIG1 *S. cerevisiae*	S H P R G K R G R K K K V V
MIG1 *K. Lactis*	D K P K G K R G R K K K S E
CREA *A. nidulans*	N N P N S R R G N K A Q H L
CREA *A. niger*	N N P N S R R R N K A Q H L
SWI5 *S. cerevisiae*	R S P R . K R G R P R K D G

b)

```
        10         20         30         40         50         60
MPQPGSSVDF SNLLNPQNNT AIPAEVSNAT ASATMASGAS LLPPMVKGAR PAAEEARQDL

        70         80         90        100        110        120
PRPYKCPLCE RAFHRLEHQT RHIRTHTGEK PHACQFPGCS KRFSRSDELT RHSRIHNNPN

       130        140        150        160        170        180
SRRGNKAQHL AAAAAAAAAN QDGSAMANNA GSMMPPPSKP ITRSAPVSQV GSPDISPPHS

       190        200        210        220        230        240
FSNYANHMRS NLSPYSRTSE RASSGMDINL LATAASQVER DESFGFRSGQ RSHHMYGPRH

       250        260        270        280        290        300
GSRGLPSLSA YAISHSMSRS HSHEDEDSYA SHRVKRSRPN SPNSTAPSSP TFSHDSLSPT

       310        320        330        340        350        360
PDHTPLATPA HSPRLKPLSP SELHLPSIRH LSLHHTPALA PMEPQAEGPN YYNPNQPHVG

       370        380        390        400        410        420
PSISDIMSRP EGAQRKLPIP QVPKVAVQDM LNPSGFTSVS SSTANSVAGG DLAERF*
```

6.6.3.1 Regulation of the alcohol utilisation regulon

Regulation of the ethanol utilisation regulon involves a double block mechanism where CREA acts to directly repress the transcription of both the pathway specific regulatory gene *alcR* and the structural genes *alcA* and *aldA* (Kulmburg *et al.*, 1993; Mathieu and Felenbok, 1994). The close proximity of the ALCR and CREA binding sites 5' of *alcR* and *alcA* have led Mathieu and Felenbok (1994) to postulate that competition between the two proteins ALCR and CREA regulates expression. Felenbok and colleagues have created a number of strains containing deletions of 5' *alcR* and 5' *alcA* sequences. In addition to this a strain which constitutively produces ALCR under the *gpdA* promoter, was constructed (Kulmburg, 1991, cited by Felenbok and Sealy-Lewis, 1994; Mathieu and Felenbok, 1994).

The *in vivo* functionality of the CREA and ALCR binding sites 5' of *alcR* has been tested by an analysis of two strains each carrying a large deletion of 5' sequences (Kulmburg *et al.*, 1992a; Kulmburg *et al.*, 1993). In the first deletion, the ALCR binding sites 5' of *alcR* were removed with the concomitant removal of the CREA **A** (and **A'**) binding sites but not the **B** or **C** sites (Fig. 3.24; Kulmburg *et al.*, 1992a). A strain containing this mutation could not be induced for *alcR* mRNA expression. This suggests that the sequences present in the 487 bp deletion which contains the *in vitro* determined ALCR binding sites, were important for ALCR mediated induction of *alcR*. Although it has been reported that this strain was normal for carbon catabolite repression (Kulmburg *et al.*, 1992a) it is not possible to say whether a strain which is not able to be induced is being repressed or if it is just simply not able to be expressed.

For the second deletion strain (*alcR*TΔA), the region 5' of *alcR* from between the **A** and **A'** CREA sites through to -1064 bp was removed (Fig. 3.24; Kulmburg *et al.*, 1993). This strain is lacking 419 bp upstream, including the **A** site, which removes the tandem repeat ALCR sites and leaves the inverted repeat ALCR sites, and the **A'**, **B**, **C1** and **C2** CREA sites, intact. Induction by ALCR was found to be retained in this strain, and derepression in the presence of glucose was observed. This result may suggest that CREA cannot prevent ALCR from

activating transcription as efficiently as it does in a wild type strain. Since only one site, **A'**, was present in the **A** region, it is possible that CREA was not able to bind to the **A** region and therefore could not prevent the ALCR protein from binding (in the presence of inducer). If the effects of an activator bound at the inverted repeats and a repressor bound at the **B** and/ or **C** sites are assumed to be additive then the overall result would be partial derepression. This was infact observed (Kulmburg *et al.*, 1993). This would also be consistent with the hypothesis (section 6.5.1) that CREA requires two binding sites for function *in vivo*. However, the sequences brought forward by such a large deletion are not known and therefore no concrete conclusions can be drawn from these experiments.

In addition, a strain was constructed where the CREA **A** site was mutated from G1, G2 to A1 and A2, but the ALCR binding sites were left intact. Northern analysis of a strain containing these point mutations showed 100 fold higher levels of *alcR* transcription in uninduced, induced and repressed/induced conditions (Mathieu and Felenbok, 1994). Only a two fold reduction in *alcR* mRNA levels for this strain was observed when glucose and inducer were present compared to a 200 fold reduction in the wild type strain. Therefore, the *alcR* **A** site is important *in vivo*. The very large loss of repression implies that most of the repression of the *alcR* gene occurs via the **A** site. However, it could be that the general increase in ALCR concentration combined with the loss of synergistic interactions between the **A**, **B** and **C** sites caused the large decrease in repression observed. Since repression was not complete, binding to the **B** and **C** sites may have resulted in some repression. This mutation leaves just one T GCG GAG site present. It is possible that derepression was due to the fact that only one site was present and/ or that now only a weak site was left. Therefore these results are consistent with the pair of sites **A** and **A'** being important *in vivo*.

The GST-ALCR binding sites 5' of *alcA* are similarly located in close proximity to the GST-CREA binding sites, although they do not overlap as was the case 5' of *alcR*. Three strains had much higher levels of *alcR* transcription. These were the strain containing *alcR* under the *gpdA*

promoter, the *alcR*TΔA strain, and the strain containing two point mutations (G1, G2 to A1 and A2) in the CREA **A** site 5' of *alcR* mentioned above (Kulmburg, 1991, cited by Felenbok and Sealy-Lewis, 1994; Kulmburg *et al.*, 1993; Mathieu and Felenbok, 1994). In all these strains repression of *alcA* by CREA, in the presence of inducer and glucose, was not complete. Mathieu and Felenbok (1994) suggest that (like the situation 5' of *alcR*) this is because in a wild type strain CREA brings about repression by competing with ALCR for binding and that in the mutant strain derepression is seen because CREA is not able to compete as effectively with the higher concentrations of ALCR protein. These researchers do not argue a role for the additional **B** site in repression. However, the fact that some repression was still observed may in part be due to binding by CREA to the **B** (and possibly the **B'**) site(s) 5' of *alcA*. It must not be forgotten that the increase in *alcR*, and *alcA* transcription in a strain overexpressing *alcR* may be due to the effect *alcR* transcription has on other processes within the cell such as increasing the intracellular concentration of metabolites or inducers. This is probably the case for (the increase in) *aldA* expression (see below).

Kulmburg *et al.*, (1992b) constructed strains which contained either a deletion from -263 to -1971 or a deletion from -731 to -1971 5' of *alcA* (see Fig. 3.25). The alcohol dehydrogenase levels of these strains and wild type under various growth conditions was compared (Kulmburg *et al.*, 1992b). The strain deleted from -263 to -1971 lacks both the **A** and **A'** CREA binding sites 5' of *alcA* and two of the three ALCR binding sites (see Fig. 3.25; Kulmburg *et al.*, 1992b). Transformants containing this mutation were found to have reduced levels of alcohol dehydrogenase activity, consistent with the removal of two thirds of the ALCR binding sites. However, alcohol dehydrogenase activity was shown to be still less in glucose grown mycelia. Since the **A** and **A'** sites were deleted in this strain these results are consistent with a role for the additional sites **B** (and possibly **B'**) in repression by CREA. The sequences brought forward by such a large deletion are not known and therefore it is possible that fortuitous CREA binding sites were brought forward. However even if this was the case it does suggest that the **A** and **A'** sites are not the only sites capable of binding CREA in order to bring about repression. These

results are also consistent with derepression being due to a loss of synergistic interactions between the **A** and **B** sites as well as being a result of the additive effects of having a repressor and an activator bound.

The presence of CREA binding sites within the small amount of available 5' *aldA* sequence suggests that direct repression by the CREA protein is likely (see Figure 6.2a). The lack of binding sites for the ALCR protein suggest that *alcR* regulates *aldA* indirectly (reviewed by Felenbok and Sealy-Lewis, 1994). The *aldA* gene product is required in a number of metabolic pathways and the complex regulation of this locus is likely to reflect this. Northern analysis of the strain containing the G1, G2 to A1 and A2 point mutations within the **A** CREA binding site 5' of *alcR*, was used to showed that with the high background of *alcR* transcription, repression of *aldA* by CREA in the presence of inducer and glucose was not complete. This is in contrast to the total repression of the *aldA* locus reported when *alcR* was constitutively expressed under the *gpdA* promoter (Kulmburg, 1991, cited by Felenbok and Sealy-Lewis, 1994; section 3.9.5.1). Since both strains produce high levels of *alcR* message it is not clear why one gave total repression of the *aldA* locus whereas the other gives partial derepression. In a wild type strain *alcR* message is expressed at a much lower level and therefore none of the strains producing *alcR* constitutively mimic the wild type situation. Furthermore CREA does not normally experience high levels of ALCR protein in the presence of glucose since in a wild type strain *alcR* transcription is tightly repressed. Therefore strains which overexpress ALCR may produce an intracellular environment quite dissimilar to what is normally present. Since no ALCR binding sites were found 5' of *aldA*, no competition between the ALCR and CREA proteins is possible. Direct repression of the *aldA* locus by CREA thus may involve negative interactions with the transcriptional machinery or competition with as yet unidentified regulators of *aldA*. Why derepression of the *aldA* locus was seen for the strain containing the G1, G2 to A1 and A2 mutations and not for the strain expressing *alcR* under the *gpdA* promoter is not immediately reconcilable. If derepression of the *aldA* locus is the true phenotype when ALCR is overexpressed then this implies that increased ALCR expression has other effects

within the cell such as increasing the intracellular concentration of metabolites, inducers or other *trans*-acting factors and that it is this which results in partial derepression.

It is possible that ALCR does not compete with CREA for binding. For example the affinity with which CREA binds DNA (especially if more than one site is present) may be far greater than the affinity with which ALCR binds DNA. There is no real evidence that CREA and ALCR fusion proteins compete for binding *in vitro*, despite the conclusions drawn by some researchers (Mathieu and Felenbok, 1994). The fact that a strain containing the *creA30* mutation is derepressed for alcohol dehydrogenase yet expresses a truncated protein containing two intact zinc fingers may suggest that straight forward competition for binding sites is not the only feature of the mechanism of repression. In chapter 5 binding studies with nuclear extracts were consistent with the CREA30 protein being able to bind the ALU148 fragment from *amdS*. Perhaps the CREA30 protein lacks a domain which allows CREA molecules to interact with each other. If repression of the *alc* genes involves synergistic interactions between CREA molecules bound at a few sites then a strain containing the *creA30* allele might be derepressed because although it may be able to prevent the *trans*-activator ALCR from binding ALCR sites which overlap with those of CREA, it lacks the synergistic interactions between sites to fully repress transcription. Alternatively, synergistic interactions may be required for high affinity binding, and a protein which can not interact with other CREA proteins may not be able to bind DNA efficiently even though the zinc finger region is stable and intact.

At higher concentrations of ALCR protein, other binding sites may be occupied by ALCR which were not able to bind GST-ALCR *in vitro,* and this may be the cause of the partial derepression of *alcR* and *alcA* that has been observed. As mentioned in section 1.1.2.3 proteins containing a Cys_6 binuclear cluster DNA binding motif such as ALCR usually bind as dimers to CCG inverted repeats spaced 6-11 bp apart. These proteins contain dimerisation domains which are required for binding. The GST-ALCR fusion protein used by Kulmburg *et al.*, (1992a,b) included only amino acids 7-58 from ALCR and did not contain any recognisable

dimerisation domain nor the putative leucine zipper domain (reviewed by Felenbok and Sealy-Lewis, 1994). Therefore the spacing between the inverted and direct repeats of GST-ALCR binding sites may have been determined by the GST moieties and not by ALCR (Todd, 1995). This may alter the type of site bound by ALCR peptides and therefore it is possible that the binding sites determined by Kulmburg *et al.*, (1992a,b) do not necessarily reflect true binding sites of the native ALCR protein, or perhaps underestimate the number of binding sites. Whether inverted or direct repeats were bound by the *Schizosaccharomyces pombe* Cys_6 binuclear cluster protein SAP1 depended on which of the two distinct and independent dimerisation domains was used (Ghazvini, 1995). Sequeval and Felenbok (1994) report that the ALCR peptide from the GST-ALCR fusion protein digested with thrombin was able to bind fragments from the 5' regions of the *alcA* and *alcR* genes. This suggests that ALCR binding is not dependent on dimerisation, and may show that ALCR binds DNA differently to other proteins in this class. Perhaps this is due to the very large number of amino acids between the third and fourth cysteine residues.

It seems probable that binding by the wild type CREA protein to at least some sites 5' of the *alc* genes will prevent binding by the transcriptional activator ALCR. However, there is no real evidence that ALCR when over expressed can compete with CREA for binding. Although some of the results presented above are consistent with a competition model for repression by CREA, some of the results point to a rather more complex combination of events. For example, there is evidence that CREA is able to repress transcription by binding to sites other than those overlapping or in close proximity to the ALCR binding sites. This is a clear indication that CREA is able to mediate repression by a competition independent method perhaps by interacting with the transcriptional machinery or by recruiting a repressor complex to carry out the repression, as is the case for MIG1. There is evidence that a single site is not as efficient at mediating repression of the *alcR* locus implying that pairs of sites are likely to be functional *in vivo*. It is surprising that the **A'** and **B'** sites 5' of *alcA* are not protected in DNase I sensitivity assays *in vitro*, however, they may be functional *in vivo*. As both loci contain two pairs of sites

there may be synergistic interactions between CREA molecules bound at each site. Sequence specific binding to the 5' region of *alcR* and *alcA* with similar sequences being recognised in both cases, and the fact that the ALCR peptide from the GST-ALCR fusion protein digested with thrombin was able to bind fragments from the 5' regions of the *alcA* and *alcR* genes that were bound *in vitro* by the GST-ALCR fusion protein, suggests that the *in vitro* determined ALCR sites are probably functional *in vivo*. It is of interest to determine whether there is any significance in the orientation of the ALCR sites (direct repeats versus inverted repeats) as this may have implications for transcriptional regulation.

6.6.3.2 Regulation of the proline utilisation cluster

The proline utilisation cluster is subject to induction, carbon catabolite and nitrogen metabolite repression. The positive regulator PRNA activates transcription in the presence of the inducer proline and *prnA* expression is not repressed by CREA whereas the structural gene for the permease, *prnB* is subject to carbon catabolite repression. Therefore this cluster is not subject to a double block mechanism of repression by CREA (Sophianopoulou *et al.*, 1993). The other genes in the cluster, *prnC* and *prnD,* have not formally been shown to be repressed by CREA since derepression of the permease may be sufficient to explain the derepression of the other genes in a prn^d20 or prn^d22 background (Sophianopoulou *et al.*, 1993). The *cis* acting prn^d20, prn^d21 and prn^d22 mutations define two adjacent binding sites as being important for repression *in vivo* (Sophianopoulou *et al.*, 1993). Binding to many sites within the controlling intergenic region by GST-CREA has been shown *in vitro* (Cubero and Scazzocchio, 1994). These additional sites may be involved in carbon catabolite repression *in vivo* as none of the *cis* acting *prn* mutations result in total derepression (Sophianopoulou *et al.*, 1993). The effects of additional CREA binding sites could be additive or synergistic. A synergistic interaction was found for the MIG1 binding sites when a hybrid protein MIG1-VP16 was assayed (section 6.6.2). If a synergistic interaction for repression between CREA binding sites did occur then the presence of many binding sites would indicate very tight control by CREA. Alternatively the presence of many binding sites could be a necessary consequence of the fact that many sites

deviate from the MIG1 consensus. Unlike CREA, MIG1 is not known to require more than one binding site for strong binding *in vitro,* and the majority of sites likely to be active *in vivo* fit the MIG1 consensus at every position.

Scazzocchio *et al.,* (1995) argue against the importance of these additional CREA binding sites *in vivo*. However, the fact that so many genes contain multiple sites and the fact that additional sites are likely to be functional 5' of the *alc* genes, suggests that these sites do play a role. These researchers hypothesise that CREA binds to a region that is a general enhancer of transcription for this regulon or that another as yet unidentified activator binds to this region, and that therefore, it is only this region which is important in regulation. This presupposes that in the absence of glucose, CREA is not able to bind DNA. This has yet to be determined. A large deletion encompassing the CREA sites, **3.1** and **3.2**, results in the *prnB* gene being repressible by ammonium but not by glucose. If the *prnB* gene is not able to be repressed at all by glucose, then this would suggest that unlike the situation 5' of *alcR,* additional CREA sites which were not deleted do not allow repression by CREA at this locus.

Like the ethanol utilisation regulon, the proline utilisation regulon is subject to specific induction by a zinc binuclear cluster containing protein, PRNA, however unlike *alcR*, *prnA* is not inducible, subject to carbon catabolite repression or autoregulated (Cazelle, 1993, cited by Scazzocchio *et al.*, 1995; Sophianopoulou *et al.*, 1993). It has been reported that the binding sites for the PRNA protein are not in close proximity to the CREA binding sites and that therefore competition between the transcriptional activator and CREA is not a feature of repression (Cazelle, 1993, cited by Scazzocchio *et al.*, 1995). This is in contrast to the situation for the *alc* genes where binding by CREA is likely to affect binding at least to some of the ALCR binding sites. Although no competition between PRNA and CREA is suggested, competition may be a feature of the regulation of the *prn* regulon if the postulated but as yet unidentified activator is found and shown to bind to the region spanning the *in vivo* active sites, **3.1** and **3.2**.

It would be of interest to determine the molecular basis of the phenotype caused by loss of amino acids carboxy terminal to residue 361 as a strain containing this mutation is derepressed for proline utilisation but not alcohol utilisation (Scazzocchio *et al.,* 1995). This finding suggests that CREA represses transcription of the *prn* and *alc* genes by different mechanisms. Evidence that AREA activates genes by different mechanisms comes from an analysis of strains containing carboxy-terminal deletions. A carboxy-terminal region of AREA which was shown by Kudla *et al.*, (1990) to be dispensable for expression of a large number of genes was found to be very important in the expression of a few genes (Stankovich, *et al.*, 1993). It is not clear how the CREA protein encoded by this allele is able to prevent binding to the proposed enhancer region, resulting in derepression of the *prn* genes but at the same time allow total repression of the *alc* genes. Rather than an enhancer sequence being present in the *prn* regulon it seems more likely that CREA represses transcription either directly or indirectly through a repressor complex, and that it is because the protein encoded by this allele lacks the ability to negatively regulate transcription or to recruit a repressor that derepression of the *prn* genes is observed. This allele encodes a protein which is expected to contain an intact DNA binding domain, and if it is assumed to contain the domain responsible for the ability of CREA molecules to interact, then this may explain the lack of derepression seen at the *alc* locus, and hence the unique phenotype of strains containing this allele. Unfortunately the effect of this mutation on *aldA* expression has not been reported.

AREA mediates nitrogen metabolite repression (section 1.3.1.4) and is only required when glucose is present and a good nitrogen source is absent. No binding studies of AREA binding to the intergenic region of *prn* have been reported, however, AREA is presumably able to bind to GATA sites present within the intergenic region to activate transcription even in the presence of active CREA (see Fig. 3.26). Competition between the CREA and AREA proteins is not likely since all GATA sites but one lie outside of the region protected in DNase I sensitivity assays. Therefore activation by AREA most likely involves interaction with the transcriptional machinery either directly or indirectly by recruiting another protein(s) to provide the activation.

6.6.3.3 Regulation of some genes involved in the utilisation of compounds via acetate metabolism

Binding to the 5' region of *amdS* and *facB* suggests that regulation of these genes at the level of transcription is directly mediated by the CREA protein. Derepression in a strain carrying a deletion of the two CREA binding sites, **A1** and **A2**, confirms the *in vivo* importance of the binding sites determined *in vitro* by DNase I sensitivity assays. Proof that CREA acts directly at the *facB* locus awaits mutation of the two CREA binding sites (**F6** and **F7**) determined by *in vitro* techniques and observation of these effects *in vivo*. It is suggested here that a double block mechanism of repression operates on the acetamidase where both the structural gene, *amdS,* and the regulatory gene, *facB,* are directly repressed by CREA. Other genes that regulate *amdS* such as *amdR* and *amdA* have not been tested to determine whether CREA affects the expression of these genes. The presence of a pair of adjacent CREA binding sites in the 5' regions of both genes (Fig. 6.2b) implies that some regulation by CREA, either positive or negative, may occur. The regulatory proteins controlling *amdS* expression co-regulate a wide variety of other genes which are involved in diverse pathways, and thus direct repression of the *amdS* gene by CREA may be the most energetically favourable mechanism.

The *amdS* gene from *A. nidulans* is also subject to nitrogen metabolite repression. Therefore, like the proline utilisation cluster, this gene provides a good model with which to study the interaction between the two global repression systems mediated by CREA and AREA. The 5' region of *amdS* contains GATA sequences (see Figure 3.21) that are potential binding sites for the AREA protein. Deletion of one of these resulted in some derepression in the presence of a good nitrogen source (M.J Hynes and M.A. Davis unpublished, cited by Davis *et al.*, 1993). This suggests that a similar situation occurs 5' of *amdS* as occurs within the *prn* intergenic region that the AREA protein can bind and activate transcription regardless of the presence or absence of the CREA protein. Most of the GATA sites lie outside of the region bound by the GST-CREA fusion protein found in this study, suggesting that competition between CREA and

AREA does not occur. Perhaps AREA is able to activate transcription by interacting more strongly with the transcriptional machinery, thus overriding repression by CREA.

The presence of CREA binding sites within the 5' region of the *facA*, *acuD* and *acuE* genes involved in acetate utilisation may suggest that CREA represses these genes directly (Fig. 6.2a). The utilisation of glucose is preferred over the utilisation of compounds metabolised via acetate and thus these genes are expected to be repressed by CREA. However, results in support of the *in vivo* importance of these sites has only been shown for the *acuD* gene (section 6.5.1). Without *in vivo* evidence to confirm that CREA directly represses these genes it is not possible to be sure that the regulation of these genes involves a "double block" method of repression as has been found for the *alc* genes, but not the proline utilisation genes. Since the CREA binding sites 5' of *facB* determined in this study to be able to bind CREA *in vitro* are a considerable distance from the start point of translation, it is possible that they are not active *in vivo*. In this case regulation of the *facA*, *acuD* and *acuE* genes may be similar to the situation for the *prn* utilisation genes where CREA represses the structural genes but not the regulatory genes.

DNA binding studies were performed using a region containing the FACB Cys_6 DNA binding domain and the putative leucine zipper dimerisation domain fused in frame with the maltose binding protein and the α chain of β-galactosidase (MBP-FacB) (Todd, 1995). Extracts containing the *E. coli* expressed fusion protein were found to bind only weakly to the 5' region of *facB* (Fig. 6.2a; Todd, 1995). This may indicate that perhaps *facB* is not directly autoregulated or that a stronger site lies further upstream. The two weak sites found do not overlap with any CREA sites, but a potential site not tested in binding studies of MBP-FacB is adjacent to the site **F2**. However, this site did not bind CREA strongly. FACB sites 5' of *facB* were determined only for the region up to 714 bp and thus it is not possible to say whether FACB binding sites are in close proximity to CREA binding sites **F6** and **F7**. An analysis of the sequence spanning the **F6** and **F7** sites does not indicate any theoretical FACB binding sites. Therefore competition between FACB and CREA is not likely 5' of *facB*.

FACB binding sites present within the 5' regions of the *facA*, *acuD, acuE* and *amdS* genes has also been determined (Fig. 6.2a; Todd, 1995). Of the three binding sites 5' of *amdS,* the strong FACB binding site overlaps with the **A4** CREA binding site (Figure 3.21). This could mean that strong activation by FACB is prevented in the presence of glucose by successful competition by the CREA protein. The FACB binding sites 5' of *acuD* were determined to be at least 50 bp from the pair of CREA binding sites likely to be important *in vivo* (Figure 6.2a). This is unlikely to be close enough to allow CREA to sterically interfere with binding by the FACB protein, however, through negative interactions with the FACB activator, CREA may be able to prevent FACB from activating. There are no FACB binding sites in close proximity to the pair of CREA binding sites 5' of *acuE* (Fig. 6.2a). There are many FACB binding sites 5' of *facA*, however, only one of these is a medium site, the rest are all weak sites. The medium site falls outside of the region present in the data base and was therefore not present in the analysis carried out here. Therefore, although the location of FACB binding sites with respect to CREA binding sites 5' of *amdS* is such that competition for binding may account for repression of transcription in the presence of inducer and glucose, for other genes regulated by *facB*, there is little evidence to support a competition model for repression by CREA.

The *facB88* mutation (discussed in section 1.3.1.1) results in greatly increased acetate induction of *amdS* but apparently normal expression of the enzymes of acetate utilisation. The binding of a GST-AmdX fusion protein to a sub-region of that protected by CREA implies that these two very closely related Cys_2His_2 finger proteins are able to recognise the same binding site(s) at least *in vitro*. Deletion studies have shown that the *facB88* gene product can lead to super activation of 5' *amdS-LacZ* expression in the presence of FACB and CREA binding sites 5' of *amdS* (Murphy, 1996). The *facB88* mutation has no effect on the enzymes of acetate utilisation, allowing the conclusion that the CREA binding sites are not correctly placed with respect to those of FACB 5' of *facB*, *acuD* or *acuE* for binding by the putative 5' FACB-AMDX hybrid protein. FACB binding sites 5' of these genes have been determined and the alternative explanation that there are no *in vivo* CREA binding sites 5' of these genes is not likely (Fig.

6.2a). A strain lacking AMDX has only slightly reduced growth on acetamide (both as a nitrogen source and a nitrogen plus carbon source) and reduced expression of 5' *amdS-LacZ* (Murphy, 1996) and thus its role in *amdS* regulation is unlikely to be major. Perhaps it is involved in a minor pathway of *amdS* regulation.

6.6.3.4 Regulation of the *creA* gene

Mobility shift assays on the 5' region of *creA* suggests that CREA acts to repress its own transcription directly. This is consistent with the finding that in the wild type, there are higher levels of *creA* mRNA in derepressed conditions (1% arabinose) compared to glucose growth conditions, and that in mutant strains there are (wild type) derepressed levels of *creA* message in both glucose and derepressing conditions (Dowzer, 1991). Higher levels of message in derepressed conditions compared with repressing conditions is not what would be expected intuitively. It means that when CREA is not required as a repressor, more mRNA is present than under conditions when CREA is required as a repressor. A switch from derepressing to repressing carbon source utilisation would bring about a rapid repression by CREA since *de novo* synthesis of the mRNA would not be required. However, it is not immediately obvious why such a rapid reaction to a change in carbon source utilisation would be advantageous. Perhaps CREA has another function under these conditions, for example a positive role in the regulation of some genes. No titration of the *creA* gene product was observed in a transformant containing in excess of 100 copies of the *amdS*$^{+}$ gene, and this is consistent with a negative autoregulatory role for the *creA* gene product (Kelly and Hynes, 1987; J.M. Kelly and M.J. Hynes pers. comm.).

There are potential AREA binding sites 5' of *creA* (Fig. 3.23), however, no results of the functionality of these sites has been reported. Potential CREA binding sites 5' of *areA* were also found. It seems unlikely that either wide domain regulatory gene is directly regulated by the other. Instead the two pathways would be expected to interact but that this is more likely to be specifically at the level of control of regulated genes.

6.7 Future work

This study has shown sequence specific binding by CREA to the 5' region of the *amdS, facB* and *creA* genes and suggests that regulation of these genes at the level of transcription is directly mediated by the CREA protein. CREA has also been shown to bind to multiple sites within the 5' regions of the *alcR, alcA* and the *prn* intergenic region *in vitro* (Kulmburg *et al.*, 1993, Cubero and Scazzocchio, 1994). CASTing experiments indicate that the targets identified within the 5' region of these genes represent the only class of sites that CREA is able to bind, and thus a mechanism whereby control by CREA depends on the type of sites bound is not likely. A number of important questions remain to be addressed. One is the identification of functional domains within CREA. The presence of domains able to activate or repress transcription can be investigated by deletion studies or by the use of reporter gene constructs. These constructs require transformation into a null allele which has recently become available. Another technique which can be used to determine whether CREA contains an activation and/ or a repression domain is to use a yeast two hybrid screen (Fields and Song, 1989). This analysis is currently underway in our laboratory.

A method of detecting CREA *in vivo* is still required, and thus further attempts to raise CREA specific antibodies will be worthwhile. In this study two CREA fusion proteins were used to generate antibodies but the sera obtained did not contain antibodies of sufficient specificity to detect CREA in nuclear extracts (chapter 5). Recently the *creA* homologue from *Trichoderma reesei* was cloned (Strauss *et al.*, 1995b), and a peptide from the linker region was used to generate antibodies. They report that specific antibodies toward CREA were obtained. It is surprising that this method worked, as at least one protein, AMDA, has been shown to contain two zinc fingers with very high sequence similarity to CREA (see Fig. 1.3), and thus antibodies directed at the linker region would be expected to also detect AMDA protein *in vivo*. However, this approach may also be effective for *A. nidulans* CREA. Another approach which can be used to detect CREA *in vivo* is to express the gene as a fusion with a small peptide for which antibodies are commercially available.

In order to carry out a complete analysis of *creA* mutant alleles, it is necessary to study the protein formed by these strains. For example, its binding, nuclear localisation, stability and protein-protein interaction properties. The generation of specific antibodies will be valuable in this analysis. Another vital question regarding CREA is whether its stability or activity is different in different growth conditions, and once again antibodies that detect CREA will be of importance in these studies. Another question for which antibodies would be useful is the question as to whether CREA functions as a monomer or a multimer, and whether this property is dependent on the growth conditions.

In order to dissect the signalling pathway, it is of interest to characterise proteins that interact with CREA, and if antibodies can be generated, then co-immunoprecipitation will be a valuable approach. It may be possible to use immunoprecipitation to isolate small quantities of these proteins from an SDS gel, and after N-terminal sequencing of these proteins, degenerate oligonucleotides can be used to clone these genes. Otherwise a yeast two hybrid screen can be used to isolate genes encoding proteins that interact with CREA.

REFERENCES

References

Adams, T.H., Boylan, M.T. and Timberlake, W.E. (1988) *brlA* is necessary and sufficient to direct conidiophore development in *Aspergillus nidulans*, *Cell*, **54**, 353-362.

Albert Hubbard, E.J., Jiang, R. and Carlson, M. (1994) Dosage-dependent modulation of glucose repression by MSN3 (STD1) in *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **14**, 1972-1978.

Alderson, A., Sabelli, P.A., Dickinson, J.R., Cole, D., Richardson, M., Kreis, M., Shewry, P.R. and Halford, N.G. (1991) Complementation of *snf1*, a mutation affecting global regulation of carbon metabolism in yeast, by a plant protein kinase cDNA, *Procedings of the National Academy of Sciences*, **88**, 8602-8605.

Alex, R., Sozeri, O., Meyer, S. and Dildrop, R. (1992) Determination of the DNA sequence recognized by the bHLH-zip domain of the N-Myc protein, *Nucleic Acids Research*, **20**, 2257-2263.

Andrianopoulos, A. and Hynes, M.J. (1988) Cloning and analysis of the positively acting regulatory gene *amdR* from *Aspergillus nidulans*, *Molecular and Cellular Biology*, **8**, 3532-3541.

Andrianopoulos, A. and Hynes, M.J. (1990) Sequence and functional analysis of the positively acting regulatory gene *amdR* from *Aspergillus nidulans*, *Molecular and Cellular Biology*, **10**, 3194-3203.

Andrianopoulos, A. and Timberlake, W.E. (1994) The *Aspergillus nidulans abaA* gene encodes a transcriptional activator that acts as a genetic switch to control development, *Molecular and Cellular Biology*, **14**, 2503-2515.

Aramayo, R. and Timberlake, W.E. (1990) Sequence and molecular structure of the *Aspergillus nidulans yA* (laccase I) gene, *Nucleic Acids Research*, **18**, 3415.

Aramayo, R. and Timberlake, W.E. (1993) The *Aspergillus nidulans yA* gene is regulated by *abaA*, *The EMBO Journal*, **12**, 2039- 2048.

Arst, H.N. (1976) Integrator gene in *Aspergillus nidulans*, *Nature*, **262**, 231-234.

Arst, H.N. (1980) Aspects of the control of gene expression in fungi, *Symposium of the Society for General Microbiology*, **31**, 131-160.

Arst, H.N. and Cove, D.J. (1973) Nitrogen metabolite repression in *Aspergillus nidulans*, *Molecular and General Genetics*, **126**, 111-141.

Arst, H.N. and MacDonald, D.W. (1975) A gene cluster in *Aspergillus nidulans* with an internally located *cis*-acting regulatory region, *Nature*, **254**, 26-31.

Arst, H.N. and Bailey, C.R. (1977) The regulation of carbon metabolism in *Aspergillus nidulans*, In J.E. Smith and J.A. Pateman (eds.) Genetics and physiology of *Aspergillus*. Academic press, Inc., N.Y., pp. 131-146.

Arst, H.N., MacDonald, D.W. and Jones, S.A. (1977) The *prn* gene cluster in *Aspergillus nidulans*, *Heredity*, **39**, 434.

Arst, H.N., MacDonald, D.W. and Jones, S.A. (1980) Regulation of proline transport in *Aspergillus nidulans*, *Journal of General Microbiology*, **116**, 285-294.

Arst, H.N., Penfold, H.A. and Bailey, C.R. (1978) Lactam utilisation in *Aspergillus nidulans*: evidence for a fourth gene under the control of the integrator gene *intA*, *Molecular and General Genetics*, **166**, 321-327.

Arst, H.N., Tollervey, D., Dowzer, C.E.A. and Kelly, J.M. (1990) An inversion truncating the *creA* gene of *Aspergillus nidulans* results in carbon catabolite derepression, *Molecular Microbiology*, **4**, 851-854.

Atkinson, P.W., King, J.A. and Hynes, M.J. (1985) Identification of the *aciA* gene controlled by the *amdA* regulatory gene in *Aspergillus nidulans*, *Current Genetics*, **10**, 133-138.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G. and Struhl, K. (1987) Current protocols in molecular biology 1987-1988, John Wiley and Sons, N.Y.

Axelrod, D.E. (1972) Kinetics of differentiation of conidiophores and conidia by colonies of *Aspergillus nidulans*, **73**, 181-184.

Bailey, C. and Arst, H.N. (1975) Carbon catabolite repression in *Aspergillus nidulans*, *European Journal of Biochemistry*, **51**, 573-577.

Baxevanis, A.D. and Vinson, C.R. (1993) Interactions of coiled coils in transcription factors: where is the specificity?, *Current Opinion in Genetics and Development*, **3**, 278-285.

Becker, D.M., Fikes, J.D. and Guarente,L. (1991) A cDNA encoding a human CCAAT-binding protein cloned by functional complementation in yeast, *Procedings of the National Academy of Sciences*, **88**, 1968-1972.

Berger, S.L., Cress, W.D., Cress, A., Triezenberg, S.J. and Guarente, L. (1990) Selective inhibition of activated but not basal transcription by the acidic activation domain of VP16: Evidence for transcriptional adapters, *Cell*, **61**, 1199-1208.

Beullens, M., Mbonyi, K., Geerts, L., Gladines, D., Detremerie, K., Jans, A.W.H. and Thevelein, J.M. (1988) Studies on the mechanism of the glucose-induced cAMP signal in glycolysis and glucose repression mutants of the yeast *Saccharomyces cerevisiae*, *European Journal of Biochemistry*, **172**, 227-231.

Blumberg, H., Eisen, A., Sledziewski, A., Bader, D. and Young, E.T. (1987) Two zinc fingers of a yeast regulatory protein shown by genetic evidence to be essential for its function, *Nature*, **328**, 443-445.

Boddy, L.M., Berges, T., Barreau, C., Vainstein, M.H., Dobson, M.J., Ballance, D.J. and Peberdy, J.F. (1993) Purification and characterisation of an *Aspergillus niger* invertase and its DNA sequence, *Current Genetics*, **24**, 60-66.

Bonnefoy, N., Copsey, J., Hynes, M.J. and Davis, M.A. (1995) Yeast proteins can activate expression through regulatory sequences of the *amdS* gene of *Aspergillus nidulans*, *Molecular and General Genetics*, **246**, 223-227.

Boulikas, T. (1993) Nuclear localization signals (NLS), *Critical Reviews in Eukaryotic Gene Expression*, **3**, 193-227.

Bowyer, P., De Lucas, J.R. and Turner, G. (1994) Regulation of the expression of the isocitrate lyase gene (*acuD*) of *Aspergillus nidulans*, *Molecular and General Genetics*, **242**, 484-489.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Analytical Biochemistry*, **72**, 248-254.

Brody, H. and Carbon, J. (1989) Electrophoretic karyotype of *Aspergillus nidulans*, *Procedings of the National Academy of Sciences*, **86**, 6260-6263.

Burger, G., Strauss, J., Scazzocchio, C. and Lang, B.F. (1991) *nirA*, the pathway-specific regulatory gene of nitrate assimilation in *Aspergillus nidulans*, encodes a putative GAL4-type zinc finger protein and contains four introns in highly conserved regions, *Molecular and Cellular Biology*, **11**, 5746-5755.

Busch, S.J. and Sassone-Corsi, P. (1990) Dimers, leucine zippers and DNA-binding domains, *Trends in Genetics*, **6**, 36-40.

Cairns, B.R., Kim,Y-J., Sayre, M.H., Laurent, B.C. and Kornberg, R.D. (1994) A multisubunit complex containing the *SWI1/ADR6*, *SWI2/SNF2*, *SWI3*, *SNF5*, and *SNF6* gene products isolated from yeast, *Procedings of the National Academy of Sciences*, **91**, 1950-1954.

Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeger, H., Lewis, W.H., Jones, C. and Houseman, D.E. (1990) Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus, *Cell*, **60**, 509-520.

Carling, D., Aguan, K., Woods, A., Verhoeven, A.J.M., Beri, R.K., Brennan, C.H., Sidebottom, C., Davison, M.D. and Scott, J. (1994) Mammalian AMP-activated protein kinase is homologous to yeast and plant protein kinases involved in the regulation of carbon metabolism, *Journal of Biological Chemistry*, **269**, 11442-11448.

Carlson, M. (1987) Regulation of sugar utilization in *Saccharomyces* species, *Journal of Bacteriology*, **169**, 4873-4877.

Carlson, M. and Botstein, D. (1982) Two differentially regulated mRNAs with different 5' ends encode secreted and intracellular forms of yeast invertase, *Cell*, **28**, 145-154.

Carlson, M. and Laurent, B.C. (1994) The SNF/SWI family of global transcriptional activators, *Current Opinion in Cell Biology*, **6**, 396-402.

Cassart, J-P., Georis, I., Ostling, J., Ronne, H. and Vandenhaute, J. (1995) The MIG1 repressor from *Kluyveromyces lactis*: cloning, sequencing and functional analysis in *Saccharomyces cerevisiae, FEBS Letters*, **371**, 191-194.

Celenza, J.L. and Carlson, M. (1984) Cloning and genetic mapping of *SNF1*, a gene required for expression of glucose repressible genes in *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **4**, 49-53.

Celenza, J.L. and Carlson, M. (1986) A yeast gene that is essential for release from glucose repression encodes a protein kinase, *Science*, **233**, 1175-1180.

Celenza, J.L. and Carlson, M. (1989) Mutational analysis of the *Saccharomyces cerevisiae* SNF1 protein kinase and evidence for functional interaction with the SNF4 protein, *Molecular and Cellular Biology*, **9**, 5034-5044.

Celenza, J.L., Eng, F.J. and Carlson, M. (1989) Molecular analysis of the *SNF4* gene of *Saccharomyces cerevisiae*: evidence for physical association of the SNF4 protein with the SNF1 protein kinase, *Molecular and Cellular Biology*, **9**, 5045-5054.

Chang, Y.C. and Timberlake, W.E. (1993) Identification of Aspergillus *brlA* response elements (BREs) by genetic selection in yeast, *Genetics*, **133**, 29-38.

Charles, I.G., Keyte, J.W., Brammar, W.J., and Hawkins, A.R. (1985) Nucleotide sequence encoding the biosynthetic dehydroquinase function of the penta-functional *AROM* locus of *Aspergillus nidulans, Nucleic Acids Research*, **13**, 8119-8128.

Charles, I.G., Keyte, J.W., Brammar, W.J., Smith, M. and Hawkins, A.R. (1986) The isolation and nucleotide sequence of the complex *AROM* locus of *Aspergillus nidulans, Nucleic Acids Research*, **14**, 2201-2203.

Chavrier, P., Zerial, M., Lemaire, P., Almendral, J., Bravo, R. and Charnay, P. (1988) A gene encoding a protein with zinc fingers is activated during G_o/G_1 transition in cultured cells, *The EMBO Journal*, **7**, 29-35.

Chavrier, P., Vesque, C., Galliot, B., Vigneron, M., Dolle, P., Duboule, D. and Charnay, P. (1990) The segment-specific gene *Krox-20* encodes a transcription factor with binding sites in the promoter region of the *Hox-1.4* gene, *The EMBO Journal*, **9**, 1209-1218.

Cho, Y., Gorina, S., Jeffrey, P.D. and Pavletich, N.P. (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations, *Science*, **265**, 346-355.

Chodosh, L.A., Olesen, J., Hahn, S., Baldwin, A.S., Guarente, L. and Sharp, P.A. (1988) A yeast and a human CCAAT-binding protein have heterologous subunits that are functionally interchangeable, *Cell*, **53**, 25-35.

Chow, C.M. and RajBhandary, U.L. (1993) Developmental regulation of the gene for formate dehydrogenase in *Neurospora crassa*, *Journal of Bacteriology*, **175**, 3703-3709.

Christy, B. and Nathans, D. (1989) DNA binding site of the growth factor-inducible protein Zif268, *Procedings of the National Academy of Sciences*, **86**, 8737-8741.

Christy, B.A., Lau, L.F. and Nathans, D. (1988) A gene activated in mouse 3T3 cells by serum growth factors encodes a protein with "zinc finger" sequences, *Procedings of the National Academy of Sciences*, **85**, 7857-7861.

Clark, K.L., Halay, E.D., Lai, E. and Burley, S.K. (1993) Co-crystal structure of the HNF-3/*fork head* DNA-recognition motif resembles histone H5, *Nature*, **364**, 412-420.

Clements, J.M. and Roberts, C.F. (1986) Transcription and processing signals in the 3-phosphoglycerate kinase (PGK) gene from *Aspergillus nidulans*, *Gene*, **44**, 97-105.

Clutterbuck, A.J. (1993) *Aspergillus nidulans*, In S.J. O'Brien (ed.) Genetic maps: Locus maps of complex genomes, 6th edition, Cold Spring Harbor Laboratory Press, Cold spring Harbor, Plainview, N.Y., pp 3.71-3.84.

Connerton, I.F., Fincham, J.R.S., Sandeman, R.A. and Hynes, M.J. (1990) Comparison and cross-species expression of the acetyl-CoA synthetase genes of the ascomycete fungi, *Aspergillus nidulans* and *Neurospora crassa*, *Molecular Microbiology*, **4**, 451-460.

Corrick, C.M., Twomey, A.P. and Hynes, M.J. (1987) The nucleotide sequence of the *amdS* gene of *Aspergillus nidulans* and the molecular characterization of 5' mutations, *Gene*, **53**, 63-71.

Cove, D.J. and Pateman, J.A. (1963) Independently segregating genetic loci concerned with nitrate reductase activity in *Aspergillus nidulans*, *Nature*, **198**, 262-263.

Covitz, P.A. Song, W. and Mitchell, A.P. (1994) Requirement for *RGR1* and *SIN4* in RME1-dependent repression in *Saccharomyces cerevisiae*, *Genetics*, **138**, 577-586.

Cubero, B. and Scazzocchio, C. (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*, *The EMBO Journal*, **13**, 407-415.

Davis, M.A. and Hynes, M.J. (1987) Complementation of *areA*- regulatory gene mutations of *Aspergillus nidulans* by the heterologous regulatory gene *nit-2* of *Neurospora crassa*, *Procedings of the National Academy of Sciences*, **84**, 3753-3757.

Davis, M.A. Cobbett, C.S. and Hynes, M.J. (1988) An *amdS-lacZ* fusion for studying gene regulation in *Aspergillus*, *Gene*, **63**, 199-212.

Davis, M.A., Kelly, J.M. and Hynes,M.J. (1993) Fungal catabolic gene regulation: molecular genetic analysis of the *amdS* gene of *Aspergillus nidulans*, *Genetica*, **90**, 133-145.

Davis, J.L., Kunisawa, R. and Thorner, J. (1992) A presumptive helicase (*MOT1* gene product) affects gene expression and is required for viability in the yeast *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **12**, 1879-1892.

Dearolf, C.R., Topol, J. and Parker, C.S. (1989) The *caudal* gene product is a direct activator of *fushi tarazu* transcription during *Drosophila* embryogenesis, *Nature*, **341**, 340-343.

de Graaff, L. and Visser, J. (1988) Structure of the *Aspergillus nidulans* pyruvate kinase gene, *Current Genetics*, **14**, 553-560.

De Lucas, J.R., Gregory, S. and Turner, G. (1994b) Analysis of the regulation of the *Aspergillus nidulans acuD* gene, encoding isocitrate lyase, by construction of a hybrid promoter, *Molecular and General Genetics*, **243**, 654-659.

De Lucas, J.R., Valenciano, S., Laborda, F. and Turner, G. (1994a) Glucose-induced inactivation of isocitrate lyase in *Aspergillus nidulans*, *Archives of Microbiology*, **162**, 409-413.

Diallinas, G., Gorfinkiel, L., Arst, H.N., Cecchetto, G. and Scazzocchio, C. (1995) Genetic and molecular characterization of a gene encoding a wide specificity purine permease of *Aspergillus nidulans* reveals a novel family of transporters conserved in prokaryotes and eukaryotes, *The Journal of Biological Chemistry*, **270**, 8610-8622.

Dowzer, C.E.A. (1991) The cloning and preliminary characterization of the *creA* gene from *Aspergillus nidulans*, Ph.D. thesis, Department of Genetics, University of Adelaide, Adelaide, Australia.

Dowzer, C.E.A. and Kelly, J.M. (1989) Cloning of the *creA* gene from *Aspergillus nidulans*: a gene involved in carbon catabolite repression, *Current Genetics*, **15**, 457-459.

Dowzer, C.E.A. and Kelly, J.M. (1991) Analysis of the *creA* gene, a regulator of carbon catabolite repression in *Aspergillus nidulans*, *Molecular and Cellular Biology*, **11**, 5701-5709.

Drysdale, M.R., Kolze, S.E. and Kelly, J.M. (1993) The *Aspergillus niger* carbon catabolite repressor encoding gene, *creA*, *Gene*, **130**, 241-245.

Dumortier, F., Arguelles, J.C. and Thevelein, J.M. (1995) Constitutive glucose-induced activation of the RAS-cAMP pathway and aberrant stationary-phase entry on glucose-containing medium in the *Saccharomyces cerevisiae* glucose-repression mutant *HEX2*, *Microbiology*, **141**, 1559-1566.

Eisen, A., Taylor, W.E., Blumberg, H. and Young, E.T. (1988) The yeast regulatory protein ADR1 binds in a zinc-dependent manner to the upstream activating sequence of *ADH2*, *Molecular and Cellular Biology*, **8**, 4552-4556.

Entian, K.-D. (1980) Genetic and biochemical evidence for hexokinase PII as a key enzyme involved in carbon catabolite repression in yeast, *Molecular and General Genetics*, **178**, 633-637.

Entian, K.-D. (1981) A carbon catabolite repression mutant of *Saccharomyces cerevisiae* with elevated hexokinase activity: evidence for regulatory control of hexokinase PII synthesis, *Molecular and General Genetics*, **184**, 278-282.

Entian, K.-D. and Barnett, J.A. (1992) Regulation of sugar utilization by *Saccharomyces cerevisiae*, *Trends in Biochemical Sciences*, **17**, 506-510.

Entian, K.-D. and Zimmermann, F.K. (1982) New genes involved in carbon catabolite repression and derepression in the yeast *Saccharomyces cerevisiae*, *Journal of Bacteriology*, **151**, 1123-1128.

Erickson, J.R. and Johnston, M. (1993) Genetic and molecular characterization of *GAL83*: Its interaction and similarities with other genes involved in glucose repression in *Saccharomyces cerevisiae, Genetics*, **135**, 655-664.

Espeso, E.A. and Penalva, M.A. (1992) Carbon catabolite repression can account for the temporal pattern of expression of a penicillin biosynthetic gene in *Aspergillus nidulans*, *Molecular Microbiology*, **6**, 1457-1465.

Espeso, E.A. and Penalva, M.A. (1994) In vitro binding of the two-finger repressor CreA to several consensus and non-consensus sites at the *ipnA* upstream region is context dependent, *FEBS Letters*, **342**, 43-48.

Espeso, E.A., Tilburn, J., Arst, H.N. and Penalva, M.A. (1993) pH regulation is a major determinant in expression of a fungal penicillin biosynthetic gene, *The EMBO Journal*, **12**, 3947-3956.

Estruch, F. and Carlson, M. (1990a) *SNF6* encodes a nuclear protein that is required for expression of many genes in *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **10**, 2544-2553.

Estruch, F. and Carlson, M. (1990b) Increased dosage of the *MSN1* gene restores invertase expression in yeast mutants defective in the SNF1 protein kinase, *Nucleic Acids Research*, **18**, 6959-6964.

Estruch, F. and Carlson, M. (1993) Two homologous zinc finger genes identified by multicopy suppression in a SNF1 protein kinase mutant of *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **13**, 3872-3881.

Everett, R.D., Barlow, P., Milner, A., Luisi, B., Orr, A., Hope, G. and Lyon, D. (1993) A novel arrangement of zinc-binding residues and secondary structure in the C_3HC_4 motif of an alpha herpes virus protein family, *Journal of Molecular Biology*, **234**, 1038-1047.

Fairall, L., Schwabe, L.C., Finch, J.T. and Rhodes, D. (1993) The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition, *Nature*, **366**, 483-487.

Federoff, H.J., Eccleshall, T.R. and Marmur, J. (1983) Carbon catabolite repression of maltase synthesis in *Saccharomyces carlsbergensis*, *Journal of Bacteriology*, **156**, 301-307.

Felenbok, B. and Sealy-Lewis, H.M. (1994) Alcohol metabolism, In Martinelli, S.D. and Kinghorn, J.R. (eds) *Aspergillus:* 50 years on, volume 29, Elsevier Science B.V. Amsterdam, The Netherlands, pp 141-179.

Felenbok, B. Sequeval, D., Mathieu, M., Sibley, S., Gwynne, D.I. and Davies, R.W. (1988) The ethanol regulon in *Aspergillus nidulans*: characterization and sequence of the positive regulatory gene *alcR*, *Gene*, **73**, 385-396.

Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions, *Nature*, **340**, 245-246.

Fillinger, S., Panozzo, C., Mathieu, M. and Felenbok, B. (1995) The basal level of transcription of the *alc* genes in the ethanol regulon in *Aspergillus nidulans* is controlled both by the specific transactivator AlcR and the general cabon catabolite repressor CreA, *FEBS Letters*, **368**, 547-550.

Flanagan, P.M., Kelleher III, R.J., Sayre, M.H., Tschochner, H. and Kornberg, R.D. (1991) A mediator required for activation of RNA polymerase II transcription *in vivo*, *Nature*, **350**, 436-438.

Flick, J.S. and Johnston, M. (1990) Two systems of glucose repression of the *GAL1* promoter in *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **10**, 4757-4769.

Fogh, R.H., Ottleben, G., Ruterjans, H., Schnarr, M., Boelens, R. and Kaptein, R. (1994) Solution structure of the LexA repressor DNA binding domain determined by ^{1}H NMR spectroscopy, *The EMBO Journal*, **13**, 3936-3944.

Forsburg, S.L. and Guarente, L. (1989) Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer, *Genes and Development*, **3**, 1166-1178.

Frederick, D.L. and Tatchell, K. (1996) The *REG2* gene of *Saccharomyces cerevisiae* encodes a type 1 protein phosphatase-binding protein that functions with REG1p and the SNF1 protein kinase to regulate growth, *Molecular and Cellular Biology*, **16**, 2922-2931.

Fu, Y.-H. and Marzluf, G.A. (1990a) *nit-2*, the major regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain, *Molecular and Cellular Biology*, **10**, 1056-1065.

Fu, Y.-H. and Marzluf, G.A. (1990b) *nit-2*, the major positive-acting nitrogen regulatory gene of *Neurospora crassa*, encodes a sequence-specific DNA-binding protein, *Procedings of the National Academy of Sciences*, **87**, 5331-5335.

Gainey, L.D.S., Connerton, I.F., Lewis, E.H., Turner, G. and Ballance, D.J. (1992) Characterization of the glyoxysomal isocitrate lyase genes of *Aspergillus nidulans* (*acuD*) and *Neurospora crassa* (*acu-3*), *Current Genetics*, **21**, 43-47.

Galas, D.J. and Schmitz, A. (1978) DNAase footprinting: a simple method for the detection of protein-DNA binding specificity, *Nucleic Acids Research*, **5**, 3157-3170.

Gancedo, J.M. (1992) Carbon catabolite repression in yeast, *European Journal of Biochemistry*, **206**, 297-313.

Garner, M.M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system, *Nucleic Acids Research*, **9**, 3047-3060.

Ghazvini, M., Ribes, V. and Arcangioli, B. (1995) The essential DNA-binding protein sap1 of Schizosaccharomyces pombe contains two independent oligomerization interfaces that dictate the relative orientation of the DNA-binding domain, M*olecular and Cellular Biology*, **15**, 4939-4946.

Glatigny, A. and Scazzocchio, C. (1995) Cloning and molecular characterization of *hxA*, the gene coding for the xanthine dehydrogenase (purine hydroxylase I) of *Aspergillus nidulans*, *Journal of Biological Chemistry*, **270**, 3534-3550.

Gong, Z. and Hew, C.L. (1994) Zinc and DNA binding properties of a novel LIM homeodomain protein Is1-2, *Biochemistry*, **33**, 15149-15158.

Gorfinkiel, L., Diallinas, G. and Scazzocchio, C. (1993) Sequence and regulation of the *uapA* gene encoding a uric acid-xanthine permease in the fungus *Aspergillus nidulans*, *The Journal of Biological Chemistry*, **268**, 23376-23381.

Griggs, D.W. and Johnston, M. (1991) Regulated expression of the *GAL4* activator gene in yeast provides a sensitive genetic switch for glucose repression, *Procedings of the National Academy of Sciences*, **88**, 8597-8601.

Gurr, S.J., Unkles, S.E. and Kinghorn, J.R. (1987) The structure and organization of nuclear genes of filamentous fungi, In: Gene structure in eukaryotic microbes, Kinghorn, J.R. (ed), IRL Press, Oxford, England, pp. 93-139.

Gwynne, D.I., Buxton, F.P., Sibley, S., Davies, R.W., Lockington, R.A., Scazzocchio, C. and Sealy-Lewis, H.M. (1987) Comparison of the *cis*-acting control regions of two coordinately controlled genes involved in ethanol utilization in *Aspergillus nidulans*, *Gene*, **51**, 205-216.

Hahn, S. and Guarente, L. (1988) Yeast HAP2 and HAP3: transcriptional activators in a heteromeric complex, *Science*, **240**, 317-321.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids, *Journal of Molecular Biology*, **166**, 557-580.

Harlow, E. and Lane, D. (1988) Antibodies, a laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Hartshorne, T.A., Blumberg, H. and Young, E.T. (1986) Sequence homology of the yeast regulatory protein ADR1 with *Xenopus* transcription factor TFIIIA, *Nature*, **320**, 283-287.

Hawkins, A.R. (1987) The complex *Arom* locus of *Aspergillus nidulans*, *Current Genetics*, **11**, 491-498.

Hayashi, S. and Scott, M.P. (1990) What determines the specificity of action of Drosophila homeodomain proteins?, *Cell*, **63**, 883-894.

Hedges, D., Proft, M. and Entian, K.-D. (1995) *CAT8*, a new zinc cluster-encoding gene necessary for derepression of gluconeogenic enzymes in the yeast *Saccharomyces cerevisiae, Molecular and Cellular Biology*, **15**, 1915-1922.

Herschbach, B.M., Arnaud, M.B. and Johnson, A.D. (1994) Transcriptional repression directed by the yeast α2 protein *in vitro*, *Nature*, **370**, 309-311.

Hintz, W.E. and Lagosky, P.A. (1993) A glucose-derepressed promoter for expression of heterologous products in the filamentous fungus *Aspergillus nidulans*, *Biotechnology*, **11**, 815-818.

Hirschhorn, J.N., Brown, S.A., Clark, C.D. and Winston, F. (1992) Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure, *Genes and Development*, **6**, 2288-2298.

Hoey, T., Warrior, R., Manak, J., and Levine, M. (1988) DNA-binding activities of the *Drosophila Melanogaster* even-skipped protein are mediated by its homeodomain and influenced by protein context, *Molecular and Cellular Biology*, **8**, 4598-4607.

Holzer, H. (1976) Catabolite inactivation in yeast, *Trends in Biochemical Sciences*, **1**, 178-181.

Hondemann, D.H.A. and Visser, J. (1994) Carbon Metabolism, In Martinelli, S.D. and Kinghorn, J.R. (eds) *Aspergillus:* 50 years on, volume 29, Elsevier Science B.V. Amsterdam, The Netherlands, pp 61-139.

Hu, Z., Nehlin, J.O., Ronne, H. and Michels, C.A. (1995) *MIG1*-dependent and *MIG1*-independent glucose regulation of *MAL* gene expression in *Saccharomyces cerevisiae*, *Current Genetics*, **28**, 258-266.

Hunter, G.D., Jones, I.G. and Sealy-Lewis, H.M. (1996) The cloning and sequencing of the *alcB* gene, coding for alcohol dehydrogenase II, in *Aspergillus nidulans, Current Genetics*, **29**, 122-129.

Hynes, M.J. (1970) Induction and repression of amidase enzymes in *Aspergillus nidulans*, *Journal of Bacteriology*, **103**, 482-487.

Hynes, M.J. (1974) Effects of ammonium, L-glutamate, and L-glutamine on nitrogen catabolism in *Aspergillus nidulans*, *Journal of Bacteriology*, **120**, 1116-1123.

Hynes, M.J. (1975a) A *cis*-dominant regulatory mutation affecting enzyme induction in the eukaryote *Aspergillus nidulans*, *Nature*, **253**, 210-212.

Hynes, M.J. (1975b) Studies on the role of the *areA* gene in the regulation of nitrogen catabolism in *Aspergillus nidulans*, *Australian Journal of Biological Sciences*, **28**, 301-313.

Hynes, M.J. (1977) Induction of the acetamidase of *Aspergillus nidulans* by acetate metabolism, *Journal of Bacteriology*, **131**, 770-775.

Hynes, M.J. (1978a) An "up-promoter" mutation affecting the acetamidase of *Aspergillus nidulans*, *Molecular and General Genetics*, **166**, 31-36.

Hynes, M.J. (1978b) Multiple independent control mechanisms affecting the acetamidase of *Aspergillus nidulans*, *Molecular and General Genetics*, **161**, 59-65.

Hynes, M.J. (1980) A mutation, adjacent to gene *amdS*, defining the site of action of positive-control gene *amdR* in *Aspergillus nidulans*, *Journal of Bacteriology*, **142**, 400-406.

Hynes, M.J. (1982) A *cis*-dominant mutation in *Aspergillus nidulans* affecting the expression of the *amdS* gene in the presence of mutations in the unlinked gene, *amdA*, *Genetics*, **102**, 139-147.

Hynes, M.J. and Kelly, J.M. (1977) Pleiotropic mutants of *Aspergillus nidulans* altered in carbon metabolism, *Molecular and General Genetics*, **150**, 193-204.

Hynes, M.J. and Pateman, J.A.J. (1970) The genetic analysis of regulation of amidase synthesis in *Aspergillus nidulans* I. Mutants able to utilize acrylamide, *Molecular and General Genetics*, **108**, 97-106.

Hynes, M.J., Corrick, C.M. and King, J.A. (1983) Isolation of genomic clones containing the *amdS* gene of *Aspergillus nidulans* and their use in the analysis of structural and regulatory mutations, *Molecular and Cellular Biology*, **3**, 1430-1439.

Hynes, M.J., Corrick, C.M., Kelly, J.M. and Littlejohn, T.G. (1988) Identification of the sites of action for regulatory genes controlling the *amdS* gene of *Aspergillus nidulans*, *Molecular and Cellular Biology*, **8**, 2589-2596.

Ish-Horowicz, D. and Burke, J.F. (1981) Rapid and efficient cosmid cloning, *Nucleic Acids Research*, **9**, 2989-2998.

Ishida, A., Asano, H., Hasegawa, M., Koseki, H., Ono, T., Yoshida, M.C., Taniguchi, M. and Kanno, M. (1993) Cloning and chromosome mapping of the human *Mel-18* gene which encodes a DNA-binding protein with a new 'RING-finger' motif, *Gene*, **129**, 249-255.

Ishizuka, H., Hanamura, A., Inada, T. and Aiba, H. (1994) Mechanism of the down-regulation of cAMP receptor protein by glucose in *Esherichia coli*: role of autoregulation of the *crp* gene, *The EMBO Journal*, **13**, 3077-3082.

Ishizuka, H., Hanamura, A., Kunimura, T. and Aiba, H. (1993) A lowered concentration of cAMP receptor protein caused by glucose is an important determinant for catabolite repression in *Esherichia coli*, *Molecular Microbiology*, **10**, 341-350.

Jackson, S.P. and Tjian, R. (1988) O-Glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation, *Cell*, **55**, 125-133.

Jacobs, G.H. (1992) Determination of the base recognition positions of zinc fingers from sequence analysis, *The EMBO Journal*, **11**, 4507-4517.

Jerlstrom, P.G., Liu, J. and Beacham, I.R. (1987) Regulation of *Escherichia coli* L-asparaginase II and L-aspartase by the *fnr* gene-product, *FEMS Microbiology Letters*, **41**, 127-130.

Johnson, P.F. and McKnight, S.L. (1989) Eukaryotic transcriptional regulatory proteins, *Annual Reviews in Biochemistry*, **58**, 799-839.

Johnston, M. and Carlson, M. (1992) Regulation of carbon and Phosphate utilization In E.W. Jones, J.R. Pringle, and J.R. Broach (eds.), Vol. II The molecular and cellular biology of the yeast *Saccharomyces*: gene regulation, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Plainview, N.Y., pp. 193-281.

Johnston, M., Flick, J.S. and Pexton, T. (1994) Multiple mechanisms provide rapid and stringent glucose repression of *GAL* gene expression in *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **14**, 3834-3841.

Johnstone, I.L., McCabe, P.C., Greaves, P., Gurr, S.J., Cole, G.E., Brow, M.A.D., Unkles, S.E., Clutterbuck, A.J., Kinghorn, J.R. and Innis, M.A. (1990) Isolation and characterisation of the *crnA-niiA-niaD* gene cluster for nitrate assimilation in *Aspergillus nidulans, Gene*, **90**, 181-192.

Jordan, S.R. and Pabo, C.O. (1988) Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions, *Science*, **242**, 893-899.

Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) Promoter-specific activation of RNA polymerase II transcription by Sp1, *Trends in Biochemical Sciences*, **11**, 20-23.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain, *Cell*, **51**, 1079-1090.

Katz, M.E. and Hynes, M.J. (1989a) Characterization of the *amdR*-controlled *lamA* and *lamB* genes of *Aspergillus nidulans*, *Genetics*, **122**, 331-339.

Katz, M.E. and Hynes, M.J. (1989b) Isolation and analysis of the acetate regulatory gene, *facB*, from *Aspergillus nidulans*, *Molecular and Cellular Biology*, **9**, 5696-5701.

Katz, M.E., Saleeba, J.A., Sapats, S.I. and Hynes, M.J. (1990) A mutation affecting *amdS* expression in *Aspergillus nidulans* contains a triplication of a *cis*-acting regulatory sequence, *Molecular and General Genetics*, **220**, 373-376.

Keleher, C.A. Redd, M.J., Schultz, J., Carlson, M. and Johnston, A.D. (1992) Ssn6-Tup1 is a general repressor of transcription in yeast, *Cell*, **68**, 709-719.

Kelly, J.M. (1980) Pleiotropic mutants of *Aspergillus nidulans* affected in carbon metabolism, Ph.D. thesis, Department of Genetics and Human Variation, La Trobe University, Melbourne, Australia.

Kelly, J.M. and Hynes, M.J. (1977) Increased and decreased sensitivity to carbon catabolite repression of enzymes of acetate metabolism in mutants of *Aspergillus nidulans*, *Molecular and General Genetics*, **156**, 87-92.

Kelly, J.M. and Hynes, M.J. (1987) Multiple copies of the *amdS* gene of *Aspergillus nidulans* cause titration of *trans*-acting regulatory proteins, *Current Genetics*, **12**, 21-31.

Kelly, J.M., Drysdale, M.R., Sealy-Lewis, H.M., Jones, I.G. and Lockington, R.A. (1990) Alcohol dehydrogenase III in *Aspergillus nidulans* is anaerobically induced and post-transcriptionally regulated, *Molecular and General Genetics*, **222**, 323-328.

Kissinger, C.R., Liu, B., Martin-Bianco, E., Kornberg, T.B. and Pabo, C.O. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions, *Cell*, **63**, 579-590.

Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. and Pabo, C.O. (1994) Crystal structure of the Oct-1 POU domain bound to an Octamer site: DNA recognition with tethered DNA-binding modules, *Cell*, **77**, 21-32.

Kolb, A., Busby, S., Buc, H., Garges, S. and Adhya, S. (1993) Transcriptional regulation by cAMP and its receptor protein, *Annual Reviews in Biochemistry*, **62**, 749-795.

Komachi, K., Redd, M.J. and Johnson, A.D. (1994) The WD repeats of Tup1 interact with the homeo domain protein α2, *Genes and Development*, **8**, 2857-2867.

Krasnow, M.A., Saffman, E.E., Kornfeld, K. and Hogness, D.S. (1989) Transcriptional activation and repression by *Ultrabithorax* proteins in cultured Drosophila cells, *Cell*, **57**, 1031-1043.

Kudla, B., Caddick, M.X., Langdom, T., Martinez-Rossi, N.M., Bennett, C.F., Sibley, S., Davies, R.W. and Arst, H.N. (1990) The regulatory gene *are*A mediating nitrogen metabolite repression in *Aspergillus nidulans*. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger, *The EMBO Journal*, **9**, 1355-1364.

Kulmburg, P., Mathieu, M., Dowzer, C., Kelly, J. and Felenbok, B. (1993) Specific binding sites in the *alcR* and *alcA* promoters of the ethanol regulon for the CREA repressor mediating carbon catabolite repression in *Aspergillus nidulans*, *Molecular Microbiology*, **7**, 847-857.

Kulmburg, P., Sequeval, D., Lenouvel, F., Mathieu, M. and Felenbok, B. (1992a) Identification of the promoter region involved in the autoregulation of the transciptional activator ALCR in *Aspergillus nidulans*, *Molecular and Cellular Biology*, **12**, 1932-1939.

Kulmburg, P., Judewicz, N., Mathieu, M., Lenouvel, F., Sequeval, D. and Felenbok, B. (1992b) Specific binding sites for the activator protein, ALCR, in the *alcA* promoter of the ethanol regulon of *Aspergillus nidulans*, *Journal of Biological Chemistry*, **267**, 21146-21153.

Kulmburg, P., Prange, T., Mathieu, M., Sequeval, D., Scazzochio, C. and Felenbok, B. (1991) Correct intron splicing generates a new type of a putative zinc-binding domain in a transcriptional activator of *Aspergillus nidulans*, *FEBS Letters*, **280**, 11-16.

Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins, *Science*, **240**, 1759-1764.

Lane, D., Prentki, P. and Chandler, M. (1992) Use of gel retardation to analyze protein-nucleic acid interactions, *Microbiological Reviews*, **56**, 509-528.

Langdon, T., Sheerins, A., Ravagnani, A., Gielkens, M., Caddick, M.X. and Arst, H.N. (1995) Mutational analysis reveals dispensability of the N-terminal region of the *Aspergillus* transcription factor mediating nitrogen metabolite repression, *Molecular Microbiology*, **17**, 877-888.

Langdon, T., Sheerins, A., Ravagnani, A., Gielkens, M., Caddick, M.X. and Arst, H.N. (1996) Mutational analysis reveals dispensability of the N-terminal region of the *Aspergillus* transcription factor mediating nitrogen metabolite repression, *Molecular Microbiology*, **20**, 239.

Laurent, B.C., Treich, I. and Carlson, M. (1993) The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation, *Genes and Development*, **7**, 583-591.

Laurent, B.C., Treitel, M.A. and Carlson, M. (1990) The SNF5 protein of *Saccharomyces cerevisiae* is a glutamine-and proline-rich transcriptional activator that affects expression of a broad spectrum of genes, *Molecular and Cellular Biology*, **10**, 5616-5625.

Lee, B.N. and Adams, T.H. (1994) The *Aspergillus nidulans fluG* gene is required for production of an extracellular developmental signal and is related to prokaryotic glutamine synthetase I, *Genes and Development*, **8**, 641-651.

Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. and Wright, P.E. (1989) Three dimentional solution structure of a single zinc finger DNA-binding domain, *Science*, **245**, 635-637.

Lesage, P., Yang, X. and Carlson, M. (1994) Analysis of the SIP3 protein identified in a two-hybrid screen for interaction with the SNF1 protein kinase, *Nucleic Acids Research*, **22**, 597-603.

Lesage, P., Yang, X. and Carlson, M. (1996) Yeast SNF1 protein kinase interacts with SIP4, a C6 zinc cluster transcriptional activator: a new role for SNF1 in the glucose response, *Molecular and Cellular Biology*, **16**, 1921-1928.

Levine, M., Rubin, G.M. and Tjian, R. (1984) Human DNA sequences homologous to a protein coding region conserved between homeotic genes of Drosophila, *Cell*, **38**, 667-673.

Li, Y., Bjorklund, S., Jiang, Y.W., Kim, Y.-J., Lane, W.S., Stillman, D.J. and Kornberg, R.D. (1995) Yeast global transcriptional regulators Sin4 and Rgr1 are components of mediator complex/ RNA polymerase II holoenzyme, *Procedings of the National Academy of Sciences*, **92**, 10864-10868.

Licht, J.D., Grossel, M.J., Figge, J. and Hansen, U.M. (1990) *Drosophila* Kruppel protein is a transcriptional repressor, *Nature*, **346**, 76-79.

Lints, R., Davis, M.A. and Hynes, M.J. (1995) The positively acting *amdA* gene of *Aspergillus nidulans* encodes a protein with two C2H2 zinc-finger motifs, *Molecular Microbiology*, **15**, 965-975.

Littlejohn, T.G. and Hynes, M.J. (1992) Analysis of the site of action of the *amdR* product for regulation of the *amdS* gene of *Aspergillus nidulans*, *Molecular and General Genetics*, **235**, 81-88.

Lockington, R.A. (1984) Molecular investigations of the ribosomal RNA genes, and the ethanol utilisation genes of *Aspergillus nidulans*, Ph.D thesis, University of Essex, Colchester, UK.

Lockington, R.A., Sealy-Lewis, H.M., Scazzocchio, C., and Davies, R.W. (1985) Cloning and characterization of the ethanol utilization regulon in *Aspergillus nidulans*, *Gene*, **33**, 137-149.

Lockington, R., Scazzocchio, C., Sequeval, D., Mathieu, M. and Felenbok, B. (1987) Regulation of *alcR*, the positive regulatory gene of the ethanol utilization regulon of *Aspergillus nidulans*, *Molecular Microbiology*, **1**, 275-281.

Lohning, C., Rosenbaum, C. and Ciriacy, M. (1993) Isolation of the *TYE2* gene reveals its identity to *SWI3* encoding a general transcription factor in *Saccharomyces cerevisiae*, *Current Genetics*, **24**, 193-199.

Lovering, R., Hanson, I., Borden, K.L.B., Martin, S., O'Reilly, N., Even, G., Rahman, D., Pappin, D.J.C., Trowsdale, J. and Freemont, P. (1993) Identification and preliminary characterization of a protein motif related to the zinc-finger, *Procedings of the National Academy of Sciences*, **90**, 2112-2116.

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R, and Sigler, P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA, *Nature*, **352**, 497-505.

Lundin, M., Nehlin, J.O. and Ronne, H. (1994) Importance of a flanking AT-rich region in target site recognition by the GC box-binding zinc finger protein MIG1, *Molecular and Cellular Biology*, **14**, 1979-1985.

Ma, H., Bloom, L.M., Walsh, C.T. and Botstein, D. (1989) The residual enzymatic phosphorylation activity of hexokinase II mutants is correlated with glucose repression in *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **9**, 5643-5649.

Magasanik, B. (1961) Catabolite repression, *Cold Spring Harbor Symposia on Quantitative Biology*, **26**, 249-256.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Mannervik, B. (1985) The isoenzymes of glutathione transferase, *Advances in Enzymology*, **57**, 357-417.

Marhoul, J.F. and Adams, T.H. (1995) Identification of developmental regulatory genes in *Aspergillus nidulans* by overexpression, *Genetics*, **139**, 537-547.

Marmorstein, R., Carey, M., Ptashne, M. and Harrison, S.C. (1992) DNA recognition by GAL4: structure of a protein-DNA complex, *Nature*, **356**, 408-414.

Marshall, M.A. and Timberlake, W.E. (1991) *Aspergillus nidulans wetA* activates spore-specific gene expression, *Molecular and Cellular Biology*, **11**, 55-62.

Martin, D.I.K. and Orkin, S.H. (1990) Transcriptional activation and DNA-binding by the erythroid factor GF-1/NF-E1/Eryf 1, *Genes and Development*, **4**, 1886-1898.

Mathieu, M. and Felenbok, B. (1994) The *Aspergillus nidulans* CREA protein mediates glucose repression of the ethanol regulon at various levels through competition with the ALCR-specific transactivator, *The EMBO Journal*, **13**, 4022-4027.

Matsumoto, K., Uno, I., Ishikawa, T. and Oshima, Y. (1983) Cyclic AMP may not be involved in catabolite repression in *Saccharomyces cerevisiae*: Evidence from mutants unable to synthesis it, *Journal of Bacteriology*, **156**, 898-900.

Mayorga, M.E. and Timberlake, W.E. (1992) The developmentally regulated *Aspergillus nidulans wA* gene encodes a polypeptide homologous to polyketide fatty acid synthases, *Molecular and General Genetics*, **235**, 205-212.

Mazon, M.J., Gancedo, J.M. and Gancedo, C. (1982a) Phosphorylation and inactivation of yeast fructose-bisphosphatase *in vivo* by glucose and by proton ionophores, a possible role for cAMP, *European Journal of Biochemistry*, **127**, 605-608.

Mazon, M.J., Gancedo, J.M. and Gancedo, C. (1982b) Inactivation of yeast fructose-1,6-bisphosphatase, *The Journal of Biological Chemistry*, **257**, 1128-1130.

McGinnis, W., Hart, C.P., Gehring, W.J. and Ruddle, F.H. (1984) Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of Drosophila, *Cell*, **38**, 675-680.

McKnight, G.L., Kato, H., Upshall, A. Parker, M.D., Saari, G. and O'Hara, P.J. (1985) Identification and molecular analysis of a third *Aspergillus nidulans* alcohol dehydrogenase gene, *The EMBO Journal*, **4**, 2093-2099.

McNabb, D.S., Xing, Y. and Guarente, L. (1995) Cloning of yeast *HAP5*: a novel subunit of a heterotrimeric complex required for CCAAT binding, *Genes and Development*, **9**, 47-58.

Merika, M. and Orkin, S.H. (1993) DNA-binding specificity of GATA family transcription factors, *Molecular and Cellular Biology*, **13**, 3999-4010.

Messing, J. (1979) A multipurpose cloning system based on single-stranded DNA bacteriophage M13, *Recombinant DNA Technology Bulletin*, **2**, 43-48.

Miller, J., McLachlan, A.D. and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes, *The EMBO Journal*, **4**, 1609-1614.

Minehart, P.L., and Magasanik, B. (1991) Sequence and expression of *GLN3*, a positive nitrogen regulatory gene of *Saccharomyces cerevisiae* encoding a protein with a putative zinc finger DNA-binding domain, *Molecular and Cellular Biology*, **11**, 6216-6228.

Mirabito, P.M., Adams, T.H. and Timberlake, W.E. (1989) Interactions of three sequentially expressed genes control temporal and spacial specificity in Aspergillus development, *Cell*, **57**, 859-868.

Mitchell, P.J. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins, *Science*, **245**, 371-378.

Moll, T., Tebb, G., Surana, U., Robitsch, H. and Nasmyth, K. (1991) The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor SWI5, *Cell*, **66**, 743-758.

Mooney, J.L. and Yager, L.N. (1990) Light is required for conididation in *Aspergillus nidulans*, *Genes and Development*, **4**, 1473-1482.

Muro, A.F., Pesce, C.G. and Kornblihtt, A.R. (1993) DNA sequencing by the chemical method: a 10 minute procedure for the G+A reaction, *Trends in Genetics*, **9**, 337-338.

Murphy, R.L. (1996) Characterisation of the *facB88* translocation of *Aspergillus nidulans*, Ph.D. thesis, Department of Genetics, University of Melbourne, Parkville, Victoria, Australia.

Murre, C., Schlonleber-McCaw,P., and Baltimore, D. (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins, *Cell*, **56**, 777-783.

Nardelli, J., Gibson, T. and Charnay, P. (1992) Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis, *Nucleic Acids Research*, **20**, 4137-4144.

Nardelli, J., Gibson, T.J., Vesque, C. and Charnay, P. (1991) Base sequence discrimination by zinc-finger DNA-binding domains, *Nature*, **349**, 175-178.

Nehlin, J. and Ronne, H. (1990) Yeast MIG1 repressor is related to the mammalian early growth response and Wilms' tumour finger proteins, *The EMBO Journal*, **9**, 2891-2898.

Nehlin, J.O., Carlberg, M. and Ronne, H. (1991) Control of yeast *GAL* genes by MIG1 repressor: a transcriptional cascade in the glucose response, *The EMBO Journal*, **10**, 3373-3377.

Nehlin, J.O., Carlberg, M. and Ronne,H. (1992) Yeast *SKO1* gene encodes a bZIP protein that binds to the CRE motif and acts as a repressor of transcription, *Nucleic Acids Research*, **20**, 5271-5278.

Neigeborn, L., Celenza, J.L. and Carlson, M. (1987) *SSN20* is an essential gene with mutant alleles that suppress defects in *SUC2* transcription in *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **7**, 672-678.

Nelson, H.C.M. (1995) Structure and funtion of DNA binding proteins, *Current Opinion in Genetics and Development*, **5**, 180-189.

Niederacher, D. and Entian, K.-D. (1991) Characterization of Hex2 protein, a negative regulatory element necessary for glucose repression in yeast, *European Journal of Biochemistry*, **200**, 311-319.

Oakley, B.R., Rinehart, J.E., Mitchell, B.L., Oakley, C.E., Carmona, C., Grey, G.L. and May, G.S. (1987) Cloning, mapping and molecular analysis of the *pyrG* (orotidine-5'-phosphate decarboxylase) gene of *Aspergillus nidulans*, *Gene*, **61**, 385-399.

O'Connell, M.J. (1990) The structure and regulation of aldehyde dehydrogenase encoding genes in *Aspergillus niger* and *Aspergillus nidulans*, Ph.D thesis, Department of Genetics, University of Adelaide, Adelaide, Australia.

O'Connell, M.J. and Kelly, J.M. (1989) Physical characterization of the aldehyde-dehydrogenase-encoding gene of *Aspergillus niger, Gene*, **84**, 173-180.

Oestreicher, N. and Scazzocchio, C. (1993) Sequence, regulation, and mutational analysis of the gene encoding urate oxidase in *Aspergillus nidulans*, *The Journal of Biological Chemistry*, **268**, 23382-23389.

Olesen, J.T., Fikes, J.D. and Guarente, L. (1991) The *Schizosaccharomyces pombe* homolog of *Saccharomyces cerevisiae* HAP2 reveals selective and stringent conservation of the small essential core protein domain, *Molecular and Cellular Biology*, **11**, 611-619.

Orejas, M., Espeso, E.A., Tilburn, J., Sarkar, S., Arst, H.N. and Penalva, M.A. (1995) Activation of the *Aspergillus* PacC transcription factor in response to alkaline ambient pH requires proteolysis of the carboxy-terminal moiety, *Genes and Development*, **9**, 1622-1632.

Orkin, S.H. (1992) GATA-binding transcription factors in hematopoietic cells, *Blood*, **80**, 575-581.

Ostling, J., Carlberg, M. and Ronne, H. (1996) Functional domains in the Mig1 repressor, *Molecular and Cellular Biology*, **16**, 753-761.

Pabo, C.O. and Sauer, R.T. (1984) Protein-DNA recognition, *Annual Reviews in Biochemistry*, **53**, 293-321.

Pan, T. and Coleman, J.E. (1990) GAL4 transcription factor is not a "zinc finger" but forms a $Zn(II)_2Cys_6$ binuclear cluster, *Procedings of the National Academy of Sciences*, **87**, 2077-2081.

Parraga, G., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T. and Klevit, R.E. (1988) Zinc-dependent structure of a single-finger domain of yeast ADR1, *Science*, **241**, 1489-1492.

Parthun, M.R. and Jaehning, J.A. (1992) A transcriptionally active form of GAL4 is phosphorylated and associated with GAL80, *Molecular and Cellular Biology*, **12**, 4981-4987.

Pastan, I. and Perlman, R. (1970) Cyclic adenosine monophosphate in bacteria, *Science*, **169**, 339-344.

Pateman, J.A., Doy, C.H., Olsen, J.E., Norris, U., Creaser, E.H. and Hynes, M.J. (1983) Regulation of alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (AldDH) in *Aspergillus nidulans*, *Procedings of the Royal Society of London series B*, **217**, 243-264.

Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1Å, *Science*, **252**, 809-817.

Perez-Esteban, B., Orejas, M., Gomez-Pardo, E. and Penalva, M.A. (1993) Molecular characterization of a fungal secondary metabolism promoter: transcription of the *Aspergillus nidulans* isopenicillin N synthetase gene is modulated by upstream negative elements, *Molecular Microbiology*, **9**, 881-895.

Perkins, A.S., Fishel, R., Jenkins, N.A. and Copeland, N.G. (1991) *Evi-1* a murine zinc finger proto-oncogene, encodes a sequence-specific DNA-binding protein, *Molecular and Cellular Biology*, **11**, 2665-2674.

Peters, D.G. and Caddick, M.X. (1994) Direct analysis of native and chimeric GATA specific DNA binding proteins from *Aspergillus nidulans*, *Nucleic Acids Research*, **22**, 5164-5172.

Peterson, C.L., Dingwall, A. and Scott, M.P. (1994) Five *SWI/SNF* gene products are components of a large multisubunit complex required for transcriptional enhancement, *Procedings of the National Academy of Sciences*, **91**, 2905-2908.

Pickett, M., Gwynne, D.I., Buxton, F.P., Elliott, R. Davies, R.W., Lockington, R.A., Scazzocchio, C. and Sealy-Lewis, H.M. (1987) Cloning and characterization of the *aldA* gene of *Aspergillus nidulans*, *Gene*, **51**, 217-226.

Pinkham, J.L. and Guarente, L. (1985) Cloning and molecular analysis of the *HAP2* locus: a global regulator of respiratory genes in *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **5**, 3410-3416.

Polakis, E.S. and Bartley, W. (1965) Changes in the enzyme activities of *Saccharomyces cerevisiae* during aerobic growth on different carbon sources, *Biochemical Journal*, **97**, 284-297.

Porter, S.D. and Smith, M. (1986) Homeo-domain homology in yeast *MATα2* is essential for repressor activity, *Nature*, **320**, 766-768.

Postma, P.W., Lengeler, J.W. and Jacobson, G.R. (1993) Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria, *Microbiological reviews*, **57**, 543-594.

Prade, R.A. and Timberlake, W.E. (1993) The *Aspergillus nidulans brlA* regulatory locus consists of overlapping transcription units that are individually required for conidiophore development, *The EMBO Journal*, **12**, 2439-2447.

Prendergast, G.C. and Ziff, E.B. (1992) A new bind for Myc, *Trends in Genetics*, **8**, 91-96.

Ptashne, M. and Gann, A.A.F. (1990) Activators and targets, *Nature*, **346**, 329-331.

Punt, P.J., Dingemanse, M.A., Jacobs-Meijsing, B.J.M., Pouwels, P.H. and van den Hondel, C.A.M.J.J. (1988) Isolation and characterization of the glyceraldehyde-3-phosphate dehydrogenase gene of *Aspergillus nidulans*, *Gene*, **69**, 49-57.

Punt, P.J., Kramer, C., Kuyvenhoven, A., Pouwels, P.H. and van den Hondel, C.A.M.J.J. (1992) An upstream activating sequence from the *Aspergillus nidulans gpdA* gene, *Gene*, **120**, 67-73.

Punt, P.J., Dingemanse, M.A., Kuyvenhoven, A., Soede, R.D.M., Pouwels, P.H. and van den Hondel, C.A.M.J.J. (1990) Functional elements in the promoter region of the *Aspergillus nidulans gpdA* gene encoding glyceraldehyde-3-phosphate dehydrogenase, *Gene*, **93**, 101-109.

Raumann, B.E., Rould, M.A., Pabo, C.O. and Sauer, R.T. (1994) DNA recognition by β-sheets in the Arc repressor-operator crystal structure, *Nature*, **367**, 754-757.

Reinert, W.R., Patel, V.B. and Giles, N.H. (1981) Genetic regulation of the *qa* gene cluster of *Neurospora crassa*: induction of *qa* messenger ribonucleic acid and dependency on *qa-1* function, *Molecular and Cellular Biology*, **1**, 829-835.

Renkawitz, R. (1990) Transcriptional repression in eukaryotes, *Trends in Genetics*, **6**, 192-197.

Richardson, I.B., Hurley, S.K. and Hynes, M.J. (1989) Cloning and molecular characterisation of the *amdR* controlled *gatA* gene of *Aspergillus nidulans*, *Molecular and General Genetics*, **217**, 118-125.

Richardson, I.B., Katz, M.E. and Hynes, M.J. (1992) Molecular characterization of the *lam* locus and sequences involved in regulation by the AmdR protein of *Aspergillus nidulans*, *Molecular and Cellular Biology*, **12**, 337-346.

Romano, A.H. and Kornberg, H.L. (1968) Regulation of sugar utilization by *Aspergillus nidulans*, *Biochimica et Biophysica Acta*, **158**, 491-493.

Romano, A.H. and Kornberg, H.L. (1969) Regulation of sugar uptake by *Aspergillus nidulans*, *Procedings of the Royal Society of London series B*, **173**, 475-490.

Rose, M., Albig,W. and Entian, K.-D. (1991) Glucose repression in *Saccharomyces cerevisiae* is directly associated with hexose phosphorylation by hexokinases PI and PII, *European Journal of Biochemistry*, **199**, 511-518.

Rose, M., Entian, K.-D., Hofmann, L., Vogel, R.F. and Mecke, D. (1988) Irreversible inactivation of *Saccharomyces cerevisiae* fructose-1,6-bisphophatase independent of protein phosphorylation at Ser^{11}, *FEBS Letters*, **241**, 55-59.

Rosenberg, U.B., Schroder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I. and Jackle, H. (1986) Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA, *Nature*, **319**, 336-339.

Roth, S.Y. (1995) Chromatin-mediated transcriptional repression in yeast, *Current Opinion in Genetics and Development*, **5**, 168-173.

Sakai, A., Shimizu,Y. and Hishinuma, F. (1988) Isolation and characterization of mutants which show an oversecretion phenotype in *Saccharomyces cerevisiae*, *Genetics*, **119**, 499-506.

Sakai, A., Shimizu,Y., Kondou, S., Chibazakura, T. and Hishinuma, F. (1990) Structure and molecular analysis of *RGR1*, a gene required for glucose repression of *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **10**, 4130-4138.

Saleeba, J.A., Cobbett, C.S. and Hynes, M.J. (1992) Characterization of the *amdA*-regulated *aciA* gene of *Aspergillus nidulans*, *Molecular and General Genetics*, **235**, 349-358.

Saluz, H.P. and Jost, J.-P. (1993) Approaches to characterize protein-DNA interactions *in vivo*, *Critical Reviews in Eukaryotic Gene Expression*, **3**, 1-29.

Sambrook, J. Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sanchez-Garcia, I. and Rabbitts, T.H. (1994) The LIM domain: a new structural motif found in zinc-finger-like proteins, *Trends in Genetics*, **10**, 315-320.

Sandeman, R.A. and Hynes, M.J. (1989) Isolation of the *facA* (acetyl-Coenzyme A synthase) and *acuE* (malate synthase) genes of *Aspergillus nidulans*, *Molecular and General Genetics*, **218**, 87-92.

Sandeman, R.A., Hynes, M.J., Fincham, J.R.S. and Connerton, I.F. (1991) Molecular organisation of the malate synthase genes of *Aspergillus nidulans* and *Neurospora crassa*, *Molecular and General Genetics*, **228**, 445-452.

Scazzocchio, C., Gavrias, V., Cubero, B., Panozzo, C., Mathieu, M. and Felenbok, B. (1995) Carbon catabolite repression in *Aspergillus nidulans*: a review, *Canadian Journal of Botany*, **73**(Suppl. 1), S160-S166.

Schmeichel, K.L. and Beckerle, M.C. (1994) The LIM domain is a modular protein-binding interface, *Cell*, **79**, 211-219.

Schork, S.M., Bee, G., Thumm, M. and Wolf, D.H. (1994) Catabolite inactivation of fructose-1,6-biphosphatase in yeast is mediated by the proteasome, *FEBS Letters*, **349**, 270-274.

Schuller, H.-J. and Entian, K.-D. (1991) Extragenic suppressors of yeast glucose derepression mutants leading to constitutive synthesis of several glucose-repressible enzymes, *Journal of Bacteriology*, **173**, 2045-2052.

Schultz, J. and Carlson, M. (1987) Molecular analysis of *SSN6*, a gene functionally related to the *SNF1* protein kinase of *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **7**, 3637-3645.

Schultz, J.E., Latter, G.I. and Matin, A. (1988) Differential regulation by cyclic AMP of starvation protein synthesis in *Escherichia coli*, *Journal of Bacteriology*, **170**, 3903-3909.

Schultz, J., Marshall-Carlson, L. and Carlson, M. (1990) The N-terminal TPR region is the functional domain of SSN6, a nuclear phosphoprotein of *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **10**, 4474-4756.

Schulz, B., Banuett, F., Dahl, M., Schlesinger, R., Schafer, W., Martin, T., Herskowitz, I. and Kahmann, R. (1990) The *b* alleles of U. maydis, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif, *Cell*, **60**, 295-306.

Scott, M.P., Tamkun, J.W. and Hartzell III, G.W. (1989) The structure and function of the homeodomain, *Biochimica et Biophysica Acta*, **989**, 25-48.

Sealy-Lewis, H.M. (1990) The identification of mutants in *Aspergillus nidulans* that lead to increased levels of ADHII, *Current Genetics*, **18**, 65-70.

Sealy-Lewis, H.M. and Lockington, R.A. (1984) Regulation of two alcohol dehydrogenases in *Aspergillus nidulans*, *Current Genetics*, **8**, 253-259.

Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins, *Procedings of the National Academy of Sciences*, **73**, 804-808.

Segall, J., Matsui, T. and Roeder, R.G. (1980) Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III, *The Journal of Biological Chemistry*, **255**, 11986-11991.

Sequeval, D. and Felenbok, B. (1994) Relationship between zinc content and DNA-binding activity of the DNA-binding motif of the transcription factor ALCR in *Aspergillus nidulans*, *Molecular and General Genetics*, **242**, 33-39.

Sharma, K.K. and Arst, H.N. (1985) The product of the regulatory gene of the proline catabolism gene cluster of *Aspergillus nidulans* is a positive-acting protein, *Current Genetics*, **9**, 299-304.

Shastry, B.S., Ng, S.-Y. and Roeder, R.G. (1982) Multiple factors involved in the transcription of class III genes in *Xenopus laevis*, *The Journal of Biological Chemistry*, **257**, 12979-12986.

Shepherd, J.C.W., McGinnis, W., Carrasco, A.E., De Robertis, E.M. and Gehring, W.J. (1984) Fly and frog homeodomains show homologies with yeast mating type regulatory proteins, *Nature*, **310**, 70-71.

Sherman, M.Y. and Goldberg, A.L. (1992) Involvement of the chaperonin dnaK in the rapid degradation of a mutant protein in *Esherichia coli*, The EMBO Journal, **11**, 71-77.

Shroff, R.A., Lockington, R.A. and Kelly, J.M., Mutational analysis of the *creA* gene involved in carbon catabolite repression in *Aspergillus nidulans*, submitted to *Molecular and General Genetics*.

Sikorski, R.S., Boguski, M.S., Goebl, M. and Hieter, P. (1990) A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis, *Cell*, **60**, 307-317.

Smith, D.B. and Johnson, K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-S-transferase, *Gene*, **67**, 31-40.

Smithies, O., Engels, W.R., Devereux, J.R., Slightom, J.L. and Shen, S. (1981) Base substitutions, length differences and DNA strand asymmetries in the human G-gamma and A-gamma fetal globin gene region, *Cell*, **26**, 345-353.

Song, W., Treich, I., Qian, N., Kuchin, S. and Carlson, M. (1996) *SSN* genes that affect transcriptional repression in *Saccharomyces cerevisiae* encode SIN4, ROX3, and SRB proteins associated with RNA polymerase II, *Molecular and Cellular Biology*, **16**, 115-120.

Sophianopoulou, V. and Scazzocchio, C. (1989) The proline transport protein of *Aspergillus nidulans* is very similar to amino acid transporters of *Saccharomyces cerevisiae*, *Molecular Microbiology*, **3**, 705-714.

Sophianopoulou, V., Suarez, T., Diallinas, G. and Scazzocchio, C. (1993) Operator derepressed mutations in the proline utilisation gene cluster of *Aspergillus nidulans*, *Molecular and General Genetics*, **236**, 209-213.

Spieth, J., Shim,Y.H., Lea, K., Conrad, R. and Blumenthal, T. (1991) *elt-1*, an embryonically expressed *Caenorhabditis elegans* gene homologous to the GATA transcription factor family, *Molecular and Cellular Biology*, **11**, 4651-4659.

Stankovich, M., Platt, A., Caddick, M.X., Langdon, T., Shaffer, P.M. and Arst, H.N. (1993) *C*-terminal truncation of the transcriptional activator encoded by *areA* in *Aspergillus nidulans* results in both loss-of-function and gain-of-function phenotypes, *Molecular Microbiology*, **7**, 81-87.

Strauss, J., Mach, R.L., Zeilinger, S. Hartler, G., Stoffler, G., Wolschek, M. and Kubicek, C.P. (1995b) Cre1, the carbon catabolite repressor protein from *Trichoderma reesei*, *FEBS Letters*, **376**, 103-107.

Strauss, J., Wolschek, M.F., Mach, R.L., Zakeri, S., Karlseder, J., Wintersberger, E. and Kubicek, C.P. (1995a) *In vivo* analysis of transcriptional regulation of the *Aspergillus nidulans creA* gene mediating glucose repression, In: The Program for the Eighteenth Fungal Genetics Conference, Asilomar Conference Centre, Pacific Grove, California, pp63.

Streatfield, S.J., Toews, S. and Roberts, C. F. (1992) Functional analysis of the expression of the 3'-phosphoglycerate kinase *pgk* gene in *Aspergillus nidulans, Molecular and General Genetics*, **233**, 231-240.

Stringer, M.A., Dean, R.A., Sewall, T.C. and Timberlake, W.E. (1991) *Rodletless*, a new *Aspergillus* developmental mutant induced by directed gene inactivation, *Genes and Development*, **5**, 1161-1171.

Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *Journal of Molecular Biology*, **189**, 113-130.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes, *Methods in Enzymology*, **185**, 60-89.

Suarez, T., de Queiroz, M.V., Oestreicher, N., and Scazzocchio, C. (1995) The sequence and binding specificity of UaY, the specific regulator of the purine utilization pathway in *Aspergillus nidulans*, suggest an evolutionary relationship with the PPR1 protein of *Saccharomyces cerevisiae*, *The EMBO Journal*, **14**, 1453-1467.

Summers, M.F., South, T.L., Kim, B. and Hare, D.R. (1990) High-resolution structure of an HIV zinc finger like domain via a new NMR-based distance geometry approach, *Biochemistry*, **29**, 329-340.

Swanson, M.S., Carlson, M. and Winston, F. (1990) *SPT6,* and essential gene that affects transcription in *Saccharomyces cerevisiae*, encodes a nuclear protein with an extremely acidic amino terminus, *Molecular and Cellular Biology*, **10**, 4935-4941.

Tagawa, M., Sakamoto, T., Kazuhiro, S., Matsubara, H., Tamura, Y., Ito, T., Okitsu, A., Imai, K. and Taniguchi, M. (1990) Expression of novel DNA-binding protein with zinc-finger structure in various tumor cells, *Journal of Biological Chemistry*, **265**, 20021-20026.

Taylor, W.E., Suruki, H.K., Lin, A.H.T., Naraghi-Arani, P., Igarashi, R.Y., Younessian, M., Katkus, P. and Vo, N.V. (1995) Designing zinc-finger ADR1 mutants with altered specificity of DNA binding to T in UAS1 sequences, *Biochemistry*, **34**, 3222-3230.

Thorley-Lawson, D.A. (1979) Characterization of cross-reacting antigens on the Epstein-Barr virus envelope and Plasma membranes of producer cells, *Cell*, **16**, 33-42.

Thukral, S.K., Eisen, A. and Young, E.T. (1991) Two monomers of yeast transcription factor ADR1 bind a palindromic sequence symmetrically to activate *ADH2* expression, *Molecular and Cellular Biology*, **11**, 1566-1577.

Thukral, S.K., Morrison, M.L. and Young, E.T. (1992) Mutations in the zinc fingers of ADR1 that change the specificity of DNA binding and transactivation, *Molecular and Cellular Biology*, **12**, 2784-2792.

Tilburn, J., Sarkar, S., Widdick, D.A., Espeso, E.A., Orejas, M., Mungroo, J., Penalva, M.A.and Arst, H.N. (1995) The *Aspergillus* PacC zinc finger transcription factor mediates regulation of both acid- and alkaline-expressed genes by ambient pH, *The EMBO Journal*, **14**, 779-790.

Timberlake, W.E. (1993) Translational triggering and feedback fixation in the control of fungal development, *The Plant Cell*, **5**, 1453-1460.

Todd, R.B. (1995) Analysis of the *facB* gene of *Aspergillus nidulans*, Ph.D. thesis, Department of Genetics, University of Melbourne, Parkville, Victoria, Australia.

Todd, R.B., Murphy, R.L., Kelly, J.M., Martin, H.M., Sharp, J.A., Davis, M.A., Katz, M.E. and Hynes, M.J. Characterization of the *Aspergillus nidulans* acetate regulatory gene *facB,* a homologue of the *CAT8* gene of *Saccharomyces cerevisiae*, submitted to *Molecular Microbiology.*

Tranque, P. Crossin, K.L., Cirelli, C., Edelman, G.M. and Mauro, V.P. (1996) Identification and characterization of a RING zinc finger gene (C-RZF) expressed in chicken embryo cells, *Procedings of the National Academy of Sciences*, **93**, 3105-3109.

Treich, I., Cairns, B.R., De Los Santos, T., Brewster, E. and Carlson, M. (1995) SNF11, a new component of the yeast SNF-SWI complex that interacts with a conserved region of SNF2, *Molecular and Cellular Biology*, **15**, 4240-4248.

Treisman, J., Harris, E., Wilson, D. and Desplan,C. (1992) The homeodomain: A new face for the helix-turn-helix?, *BioEssays*, **14**, 145-150.

Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. and Desplan, C. (1989) A single amino acid can determine the DNA binding specificty of homeodomain proteins, *Cell*, **59**, 553-562.

Treitel, M.A. and Carlson, M. (1995) Repression by SSN6-TUP1 is directed by MIG1, a repressor/ activator protein, *Procedings of the National Academy of Sciences*, **92**, 3132-3136.

Triezenberg, S.J (1995) Structure and function of transcriptional activation domains, *Current Opinion in Genetics and Development*, **5**, 190-196.

Tu, J.L. and Carlson, M. (1995) REG1 binds to protein phophatase type 1 and regulates glucose repression in *Saccharomyces cerevisiae*, *The EMBO Journal*, **14**, 5939-5946.

Tzamarias, D. and Struhl, K. (1994) Functional dissection of the yeast Cyc8-Tup1 transcriptional co-repressor complex, *Nature*, **369**, 758-761.

Tzamarias, D. and Struhl, K. (1995) Distinct TPR motifs of Cyc8 are involved in recruiting the Cyc8-Tup1 corepressor complex to differentially regulated promoters, *Genes and Development*, **9**, 821-831.

Unkles, S.E., Hawker, K.L., Grieve, C., Campbell, E.I., Montague, P. and Kinghorn, J.R. (1991a) *crnA* encodes a nitrate transporter in *Aspergillus nidulans*, *Procedings of the National Academy of Sciences*, **88**, 204-208.

Unkles, S.E., Hawker, K.L., Grieve, C., Campbell, E.I., Montague, P. and Kinghorn, J.R. (1991b) *crnA* encodes a nitrate transporter in *Aspergillus nidulans*, *Procedings of the National Academy of Sciences*, **88**, 4564.

Unkles, S.E., Campbell, E.I., Punt, P.J., Hawker, K.L., Contreras, R., Hawkins, A.R., van den Hondel, C.A.M.J.J. and Kinghorn, J.R. (1992) The *Aspergillus niger niaD* gene encoding nitrate reductase: upstream nucleotide and amino acid sequence comparisons, *Gene* , **111**, 149-155.

Uno, I., Matsumoto, K., Adachi, K. and Ishikawa, T. (1983) Genetic and biochemical evidence that trehalase is a substrate of cAMP-dependent protein kinase in yeast, *Journal of Biological Chemistry*, **258**, 10867-10872.

Upshall, A. (1986) Genetic and molecular characterization of *arg*B+ transformants of *Aspergillus nidulans*, *Current Genetics*, **10**, 593-599.

Upshall, A., Gilbert, T., Saari, G., O'Hara, P.J., Weglenski, P., Berse, B., Miller, K. and Timberlake, W.E. (1986) Molecular analysis of the *arg*B gene of *Aspergillus nidulans, Molecular and General Genetics*, **204**, 349-354.

Vallier, L.G. and Carlson, M. (1994) Synergistic release from glucose repression by *mig1* and *ssn* mutations in *Saccharomyces cerevisiae*, *Genetics*, **137**, 49-54.

Van Aelst, L., Jans, A.W.H. and Thevelein, J.M. (1991) Involvement of the *CDC25* gene product in the signal trasmission pathway of the glucose-induced RAS-mediated cAMP signal in the yeast *Saccharomyces cerevisiae*, *Journal of General Microbiology*, **137**, 341-349.

van Heeswijck, R. and Hynes, M.J. (1991) The *amdR* product and a CCAAT-binding factor bind to adjacent, possibly overlapping DNA sequences in the promoter region of the *Aspergillus nidulans amdS* gene, *Nucleic Acids Research*, **19**, 2655-2660.

Vojtek, A.B. and Fraenkel, D.G. (1990) Phosphorylation of yeast hexokinases, *European Journal of Biochemistry*, **190**, 371-375.

Wang, X.W., Hynes, M.J. and Davis, M.A. (1992) Structural and functional analysis of the *amdR* regulatory gene of *Aspergillus oryzae*, *Gene*, **122**, 147-154.

Weber, I.T. and Steitz, T.A. (1987) Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution, *Journal of Molecular Biology*, **198**, 311-326.

Williams, F.E. and Trumbly, R.J. (1990) Characterization of *TUP1*, a mediator of glucose repression in *Saccharomyces cerevisiae*, *Molecular and Cellular Biology*, **10**, 6500-6511.

Williams, F.E., Varanasi, U. and Trumbly, R.J. (1991) The CYC8 and TUP1 proteins involved in glucose repression in *Saccharomyces cerevisiae* are associated in a protein complex, *Molecular and Cellular Biology*, **11**, 3307-3316.

Wills, C. (1990) Regulation of sugar and ethanol metabolism in *Saccharomyces cerevisiae*, *Critical Reviews in Biochemistry and Molecular Biology*, **25**, 245-280.

Wilson, L.J., Carmona, C.L. and Ward, M. (1988) Sequence of the *Aspergillus niger pyrG* gene, *Nucleic Acids Research*, **16**, 2339.

Wingender-Drissen, R. and Becker, J.U. (1983) Regulation of yeast phosphorylase by phosphorylase kinase and cAMP-dependent protein kinase, *FEBS Letters*, **163**, 33-36.

Winston, F. and Carlson, M. (1992) Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection, *Trends in Genetics*, **8**, 387-391.

Woods, A., Munday, M.R., Scott, J., Yang, X., Carlson, M. and Carling, D. (1994) Yeast SNF1 is functionally related to mammalian AMP-activated protein kinase and regulates acetyl-CoA carboxylase *in vivo*, *The Journal of Biological Chemistry*, **269**, 19509-19515.

Wright, W.E. and Funk, W.D. (1993) CASTing for multicomponent DNA-binding complexes, *Trends in Biochemical Sciences*, **18**, 77-80.

Xing, Y., Fikes, J.D. and Guarente, L. (1993) Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain, *The EMBO Journal*, **12**, 4647-4655.

Yang, H.-Y. and Evans, T. (1992) Distinct roles for the two cGATA-1 finger domains, *Molecular and Cellular Biology*, **12**, 4562-4570.

Yang, X., Hubbard, E.J.A. and Carlson, M. (1992) A protein kinase substrate identified by the two hybrid system, *Science*, **257**, 680-682.

Yang, X., Jiang, R. and Carlson, M. (1994) A family of proteins containing a conserved domain that mediates interaction with the yeast SNF1 protein kinase complex, *The EMBO Journal*, **13**, 5878-5886.

Yoshinaga, S.K., Peterson, C.L., Herskowitz, I. and Yamamoto, K.R. (1992) Roles of SWI1, SWI2, and SWI3 proteins for transcriptional enhancement by steroid receptors, *Science*, **258**, 1598-1604.

Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W. and Schreiber, S.L. (1994) Structural basis for the binding of proline-rich peptides to SH3 domains, *Cell*, **76**, 933-945.

Zawel, L. and Reinberg, D. (1992) Advances in RNA polymerase II transcription, *Current Opinion in Cell Biology*, **4**, 488-495.

APPENDICES

Appendix A Results of the CASTing experiment

Name	Sequence	Affinity	Mobility[a]
L100	AACGAATTGAACGCCATATACAACGCGGGTGTGGGGTTACACCATTGA	+	HMC
L101	AACGAATTGAACGCCGCGACGATCATATTTGTGGGGTTACACCATTGA	+	HMC
L102	AACGAATTGAACGCGCGGTCGTGACATCATGGCGGGTTACACCATTGA	-	nd
L103	AACGAATTGAACGCCCATGTTATATATCCTGTGGGGTTACACCATTGA	+	HMC
L104	AACGAATTGAACGCGCCCCTTCTCATAGACTGCGGGTTACACCATTGA	-	nd
L105	AACGAATTGAACGCCACGAGCGGGTATAGCCTGGGGTTACACCATTGA	-/+	HMC
L106	AACGAATTGAACGCCATCCGCATTTAGTCTGTGGGGTTACACCATTGA	+	HMC
L107	AACGAATTGAACGCCACCGGGTCAAATTGTGGGGGGTTACACCATTGA	+	HMC
L108	AACGAATTGAACGCGGGGTGCGCTACGTATGGCGGGTTACACCATTGA	++	LMC
L109	AACGAATTGAACGCCCAACGCTGATCCATGCCGGGGTTACACCATTGA	-	nd
L110	AACGAATTGAACGCCATTGTAGAGCACATCCCGGGGTTACACCATTGA	+	HMC
L111	AACGAATTGAACGCCGGATGGGGAAGTAGCGTGGGTTACACCATTGA	-/+	HMC
L112	AACGAATTGAACGCCCCAGTTCGACCTTCCCCGGGGTTACACCATTGA	+	HMC
L113	NO INSERT		
L114	AACGAATTGAACGCCACGTATATGGTCCGCGCGGGGTTACACCATTGA	+	HMC
L115	AACGAATTGAACGCCCGACTTTTGTTCATCGTGGGGTTACACCATTGA	-/+	HMC
L116	AACGAATTGAACGCGGCCCTCATGTAGATCCCGGGGTTACACCATTGA	-/+	HMC
L117	NO INSERT		
L118	AACGAATTGAACGCCCGAGGTATACTTAATGCGGGGTTACACCATTGA	+	HMC
L119	AACGAATTGAACGCGGCAGGGGTTGCTTATGCGGGGTTACACCATTGA	+	HMC
L120	NO INSERT		
L121	AACGAATTGAACGCCACATTTGTCCCTCGCCGCGGGTTACACCATTGA	-	nd
L122	AACGAATTGAACGCGGCCGAGATGGTCATAGTGGGGTTACACCATTGA	-	nd

L123	AACGAATTGAACGCCACAACTTGCTATTTTGTGGGGTTACACCATTGA	+	HMC
L124	AACGAATTGAACGCGGAGGAGATTGTGGGGGATGGGTTACACCATTGA	++	LMC
L125	AACGAATTGAACGCCATAACAGAAGATTACGCGGGGTTACACCATTGA	+	HMC
L126	AACGAATTGAACGCCACGTTACCATATCGGCGGGGTTACACCATTGA	+	HMC
L127	AACGAATTGAACGCCACTGTGGATCCGCGTGTGGGGTTACACCATTGA	+	HMC
L128	AACGAATTGAACGCCATGCTCGCAAACTAAGTGGGGTTACACCATTGA	+	HMC
L129	AACGAATTGAACGCCACAAAGAGCAATCATGCGGGGTTACACCATTGA	+	HMC
L130	NO PRIMER		
L131	AACGAATTGAACGCGCCTTCTCCTTGTTCCGTGGGGTTACACCATTAA	+	HMC
L132	AACGAATTGAACGCCATGTGGGGGTCCTTGGCAGGGTTACACCATTGA	+	HMC
L133	AACGAATTGAACGCTGGCGGGGGTCTCACAGTGGGGTTACACCATTGA	++	LMC
L134	AACGAATTGAACGCCCCACATTTTTTCACGTGCGGGTTACACCATTGA	++	HMC
L135	NO PRIMER		
L136	NO INSERT		
L137	AACGAATTGAACGCCCCAGACCCCGTACTTGCCAGGTTACACCATTGA	+	LMC
L138	AACGAATTGAACGCGGTGCCATGGGACATTTGCAGGTTACACCATTGA	-	nd
L139	AACGAATTGAACGCGGGTGGTCGTGACACGTTGGGGTTACACCATTGA	+	LMC
L140	AACGAATTGAACGCCGCCATATGTTTTTTCCGGGGTTACACCATTGA	+	HMC
L141	AACGAATTGAACGCCACGCAACCGCTAGATGTGGGGTTACACCATTGA	+	HMC
L142	AACGAATTGAACGCCATGACAATTCTGGTTGCGGGGTTACACCATTGA	+	HMC
L144	AACGAATTGAACGCGGGGACAGTTGCTTTTGTGGGGTTACACCATTGA	++	LMC
L145	AACGAATTGAACGCGGGGGTATGCGGTGTGCTGGGGTTACACCATTGA	++	LMC
L147	AACGAATTGAACGCCATGGAGGGGGACTTGCGGGGGTTACACCATTGA	+	LMC
L148	AACGAATTGAACGCCGTGCGGACCGTGGCTGTGGGGTTACACCATTGA	+	HMC
L149	AACGAATTGAACGCCGACATGGAGGGAGCTGGGGGGTTACACCATTGA	-	nd
L150	AACGAATTGAACGCGCCCAACTTAATATTCCTGGGGTTACACCATTGA	+	HMC

S100	AACGAATTGAACGCGGAGAGCATATATGGCCGGGGGTTACACCATTGA	+	LMC
S101	AACGAATTGAACGCTACTATAGACGCGGTATAGGGGTTACACCATTGA	-	nd
S102	AACGAATTGAACGCGGGGTGCTCTCGTTGTGCGGGGTTACACCATTGA	++	LMC
S103	AACGAATTGAACGCCACGTTACTATAGCTTGCGGGGTTACACCATTGA	+	HMC
S104	AACGAATTGAACGCGGAGTACGCCTACAGCAGCGGGTTACACCATTGA	-	nd
S105	AACGAATTGAACGCCGTCGAGGTGAGGGATGTGGGGTTACACCATTGA	+	HMC
S106	AACGAATTGAACGCCAGGAGAACGCATTCTGTGGGGTTACACCATTGA	+	HMC
S107	AACGAATTGAACGCCGCGAGGTCTGCCCATGTGGGGTTACACCATTGA	+	HMC
S108	AACGAATTGAACGCCGGCATTGGACTGGGTGTGGGGTTACACCATTGA	+	HMC
S109	AACGAATTGAACGCCACTGATCGTAAGTGTGTGGGGTTACACCATTGA	+	HMC
S110	AACGAATTGAACGCCATCTCTAGTGAGTTGCGGGGGTTACACCATTGA	+	HMC
S111	AACGAATTGAACGCGCAAAAGTCCAGCTGTGTGGGGTTACACCATTGA	+	HMC
S112	AACGAATTGAACGCCACAATTGCAGTTTGGGTGGGGTTACACCATTGA	+	HMC
S113	AACGAATTGAACGCGGGTAAGATTGTGATCTGGGGTTACACCATTGA	+	HMC
S114	NO INSERT		
S115	AACGAATTGAACGCGGGGGCACACTTTTGCAAGGGGTTACACCATTGA	++	LMC
S117	AACGAATTGAACGCGGGTGAAGTATCATTCGCGGGTTACACCATTGA	-/+	LMC
S118	AACGAATTGAACGCACCTGATGCAACTTTCGCGGGGTTACACCATTGA	+	HMC
S119	AACGAATTGAACGCGGGGGTCGCCTGAGTTGCATGGTTACACCATTGA	++	LMC
S120	AACGAATTGAACGCGGACTATGCTGAGATGTGGGGGTTACACCATTGA	+	HMC
S121	AACGAATTGAACGCCACTTGATGTGAGACTGTGGGGTTACACCATTGA	+	LMC
S122	AACGAATTGAACGCGGAGGGCGCTGTACGTGTGGGGTTACACCATTGA	++	LMC
S123	AACGAATTGAACGCCACGCGGGGTATTAGGACACGGTTACACCATTGA	+	HMC
S124	AACGAATTGAACGCGGGAGATGACGTCTGTGTGGGGTTACACCATTGA	+	HMC
S125	AACGAATTGAACGCACGTTGGGGCGAGTATGTGGGGTTACACCATTGA	++	LMC
S126	AACGAATTGAACGCGGGGAAATGGTATAATGGCGGGTTACACCATTGA	++	LMC

S127	AACGAATTGAACGCGGTATGTGGGTTTTATGTGGGGTTACACCATTGA	++	LMC
S128	AACGAATTGAACGCGGGGGGGGGCGCGTCTCGTGGGTTACACCATTGA	+	LMC
S129	AACGAATTGAACGCGGCCCTAGTACCCTATGTGGGGTTACACCATTGA	+	HMC
S130	AACGAATTGAACGCGGGGCAGTTACTCATGGCGGGTTACACCATTGA	++	LMC
S131	AACGAATTGAACGCCCGTACGTTGCCCATTGCGGGGTTACACCATTGA	+	HMC
S132	AACGAATTGAACGCGGAAAGGTTACTCACTGTGGGGTTACACCATTGA	+	HMC
S133	AACGAATTGAACGCGGGCGACTACGCTTATGTGGGGTTACACCATTGA	++	LMC
S134	AACGAATTGAACGCCGTTAACTTCATGAATGTGGGGTTACACCATTGA	+	HMC
S135	AACGAATTGAACGCGGAGTACGAGGGGATTGCGGGGTTACACCATTGA	++	LMC
S136	AACGAATTGAACGCCAACGAGTCACTCCCTGCTAGGTTACACCATTGA	-	nd
S137	AACGAATTGAACGCCATCCCGGTAATATTCGTGGGGTTACACCATTGA	+	HMC
S138	AACGAATTGAACGCCACGTTTTCTGGGTGCCCGGGGTTACACCATTGA	-	nd
S139	AACGAATTGAACGCCGCACATACTAGATTGGTGGGGTTACACCATTGA	+	HMC
S140	AACGAATTGAACGCGGGGGGGATTTTGTTTGGCGGGTTACACCATTGA	++	LMC
S141	AACGAATTGAACGCCACCGAATAGATTTGTGTGGGGTTACACCATTGA	+	HMC
S142	AACGAATTGAACGCTGATAAAAATAGTGGTGCGGGGTTACACCATTGA	+	HMC
S143	AACGAATTGAACGCCGGTATCTTCGGAATTGTGGGGTTACACCATTGA	+	HMC

[a] HMC= higher mobility complex

LMC= lower mobility complex

nd= not determined

Appendix B

Table B.1 Base composition of "S" oligonucleotides forming a HMC and binding with scores of + or ++

Bases	Position relative to the 3' G of the 4G string																		
	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3
A	2	6	4	7	11	10	7	3	6	8	9	3	5	4	7					
C	16	5	8	5	5	2	3	2	6	5	6	4	7	3	2	3	1	6		
G	4	12	8	4	5	3	6	9	3	5	4	9	5	6	7	2	22		23	23
T	1		3	7	2	8	7	9	8	5	4	7	6	10	7	18		17		
A+T	3	6	7	14	13	18	14	12	14	13	13	10	11	14	14	18	0	17	0	0
C+G	19	17	16	9	10	5	9	11	9	10	10	13	12	9	9	5	23	6	23	23

Table B.2 Base composition of "L" oligonucleotides forming a HMC and binding with scores of + or ++

Bases	Position relative to the 3' G of the 4G string																		
	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3
A	1	11	2	7	8	6	4	11	5	7	6	9	6	4	7	1				
C	18	9	9	5	6	7	4	2	5	6	7	2	4	5	5	5	3	8		
G	3	2	4	6	5	3	6	6	4	4	4	2	5	4	3	1	19		22	22
T			7	4	3	6	8	3	8	5	5	9	7	9	7	15		14		
A+T	1	11	9	11	11	12	12	14	13	12	11	18	13	13	14	16	0	14	0	0
C+G	21	11	13	11	11	10	10	8	9	10	11	4	9	9	8	6	22	8	22	22

Table B.3 Base composition of all oligonucleotides forming a HMC that bind with scores of + or ++

Bases	Position relative to the 3' G of the 4G string																		
	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3
A	3	17	6	14	19	16	11	14	11	15	15	12	11	8	14	1				
C	32	13	17	10	11	9	7	4	11	11	13	6	11	8	7	8	4	14		
G	6	14	12	10	10	6	12	15	7	9	8	11	10	10	10	3	41		45	45
T	1		10	11	5	14	15	12	16	10	9	16	13	19	14	33		31		
A+T	4	17	16	25	24	30	26	26	27	25	24	28	24	27	28	34	0	31	0	0
C+G	38	27	29	20	21	15	19	19	18	20	21	17	21	18	17	11	45	14	45	45