THE CLONING AND PRELIMINARY CHARACTERIZATION OF THE creA GENE FROM Aspergillus nidulans.

by

Celia E. A. Dowzer, B.sc.(Hons.)

Department of Genetics, University of Adelaide, South Australia.

TABLE OF CONTENTS.

Abstract
Declaration
Acknowledgements
Publications
Abbreviations
List of tables
List of figures

CHAPTER ONE: INTRODUCTION
1.1 MOLECULAR MECHANISMS OF GENE REGULATION IN EUKARYOTES
  1.1.1 DNA binding motifs
  1.1.2 Activation domains
  1.1.3 Negative control systems
  1.1.4 The regulation of synthesis and activity of transcription factors
1.2 CARBON CATABOLITE REPRESSION IN Escherichia coli
1.3 CARBON CATABOLITE REPRESSION IN Saccharomyces
  1.3.1 Regulation of the galactose regulon of S. cerevisiae
  1.3.2 Sucrose utilization in S. cerevisiae
1.4 GLOBAL GENE REGULATION IN A. nidulans
1.5 CARBON CATABOLITE REPRESSION IN A. nidulans
  1.5.1 Selection of mutations affecting carbon
catabolite repression in A. nidulans

1.5.2 The phenotype of creA mutants

1.5.3 The phenotype of creB and creC mutants

1.5.4 Proposed roles for creA, creB, creC and cre-34 gene products

1.5.5 Molecular studies of regulation by carbon catabolite repression in A. nidulans

1.6 AIMS OF THIS STUDY

CHAPTER TWO: MATERIALS AND METHODS

2.1 MATERIALS

2.2 BUFFERS AND STOCK SOLUTIONS

2.2.1 1 X Reaction buffers

2.2.2 Stock solutions

2.3 MEDIA

2.3.1 Bacterial and yeast growth media

2.3.2 Aspergillus growth media

2.4 E. coli STRAINS AND BACTERIOPHAGE

2.5 FUNGAL STRAINS

2.6 PLASMIDS

2.7 METHODS

2.7.1 Genetic manipulation and growth testing of Aspergillus

2.7.2 Transformation of Aspergillus

2.7.3 Isolation of nucleic acids

2.7.4 General recombinant DNA methods

2.7.5 Gel electrophoresis
2.7.6 $^{32}$P Labelling of DNA
2.7.7 Nucleic acid blotting and hybridization conditions
2.7.8 Quantification of DNA copy number and mRNA levels
2.7.9 DNA sequencing
2.7.10 Primer extension analysis

CHAPTER THREE: THE CLONING OF creA FROM Aspergillus nidulans

3.1 THE CLONING OF creA
3.1.1 The construction of a wildtype genomic library from A. nidulans
3.1.2 The identification of a non-revertable creA- allele
3.1.3 Complementation of the creA204 mutation
3.1.4 Rescue of transforming sequences
3.1.5 The localization of creA within pANCl

3.2 ANALYSIS OF CreA+ TRANSFORMANTS
3.3 EXPRESSION OF creA IN A. nidulans
3.4 SUMMARY AND CONCLUSIONS

CHAPTER FOUR: ANALYSIS OF creAd-30 AND THE PHENOTYPE OF A creA NULL ALLELE

4.1 MOLECULAR LOCALIZATION OF THE INVERSION BREAKPOINT IN creAd-30
4.2 CONSTRUCTION OF A creA DELETION STRAIN
4.2.1 The isolation of a larger creA genomic clone 103
4.2.2 The construction of a creA replacement vector 104
4.2.3 The creation of a creA gene replacement strain of A. nidulans 106
4.2.4 Vegetative versus conidial lethality of the creA null phenotype 115
4.2.5 Rescuing the ability of TC6330 to form haploids containing chromosome I having the gene replacement 117

4.3 SEQUENCES HOMOLOGOUS TO creA FROM A. nidulans 119
4.4 SUMMARY AND CONCLUSIONS 119

CHAPTER FIVE: THE PHYSICAL CHARACTERIZATION OF THE Aspergillus nidulans creA GENE 122

5.1 THE PHYSICAL ANALYSIS OF creA 122
5.1.1 Nucleotide sequence analysis of creA and characterization of the coding region 122
5.1.2 The 5' region of the creA gene 126
5.1.3 The 3' region of the creA gene 133

5.2 ANALYSIS OF THE PUTATIVE CreA PROTEIN 133
5.2.1 The zinc finger motif 134
5.2.2 An alanine rich region 141
5.2.3 The S/TPXX motif 149
5.2.4 PEST sequences 149
5.2.5 Other general features of the CreA protein 150

5.3 SUMMARY 151
ABSTRACT

Previous genetic analyses have led to the suggestion that the product of the creA gene from Aspergillus nidulans acts as a negatively acting regulatory protein in carbon catabolite repression. This suggestion was based on the finding that common creA mutant alleles lead to the derepression of synthesis, in the presence of D-glucose, of many enzymes for the utilization of alternative carbon sources which are repressed during growth on D-glucose in wildtype A. nidulans, and these mutant alleles are recessive to the wildtype allele in diploids.

The creA region has been cloned by complementation of a mutant allele by transformation with a genomic library followed by marker rescue techniques. The creA gene encodes a transcript of approximately 2kb in length which is constitutively expressed. The nucleotide sequence of the creA genomic region and number of cDNA clones has been determined. The 390 amino acid sequence of the putative creA protein has a number of characteristic features found in regulatory and DNA binding proteins including two C2H2 zinc finger motifs and several alanine rich regions. Although no overall sequence similarities have been found to other proteins of known function, regions of similarity, possibly representing functional domains, have been identified and their significance is discussed.

All creA mutant alleles analysed produce a creA transcript. One creA allele, creA<sup>d</sup>-30 contains a translocation breakpoint within the
creA gene and has two transcripts. Thus none of the \textit{in vivo} isolated mutants tested were null alleles at the level of mRNA. A plasmid construct containing the \textit{riboB} gene from \textit{A. nidulans} flanked by sequences up and downstream of the \textit{creA} gene was used to replace one of the wildtype \textit{creA} genes in a diploid strain. Haploidization analysis of this diploid demonstrates that total loss of \textit{creA} function is lethal in vegetative cells.
DECLARATION.

I declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and to the best of my knowledge and belief contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

I give my consent for this thesis to be made available for photocopying and loan.

Celia E. A. Dowzer.
ACKNOWLEDGEMENTS.

I am extremely grateful for the excellent supervision given by Dr. Joan Kelly during my candidature in this department. I would like to thank Joan on a personal basis for her friendship, along with the encouragement and enthusiasm she has shown for this project.

I appreciate the valuable discussions, comments and suggestions on written and/or experimental work provided by Dr. Robin Lockington, Dr. Meryl Davis and Prof. Michael Hynes and the interest they have shown in the progress of this work.

I am indebted to Dr. Matthew O'Connell for his friendship and the unfailing encouragement he has given. I also appreciate very much his advice on experimental techniques and written work and the time he has spent helping with computer analyses.

Dr. Dara Whisson provided many suggestions on experimental methods and has also been a good friend over the past few years. I would also like to thank Angela Binns and Michelle Adamou for their technical help in lab109. I wish to thank Doug Anderson and Ulrik John for their time and effort in proof-reading this thesis.

I acknowledge the financial support of a Commonwealth Postgraduate Research Award and an Australian Postgraduate Priority Research Award.
PUBLICATIONS.

The work presented in this thesis has been submitted for publication or published under the following titles:


ABBREVIATIONS.

A: Adenosine; bp: base pair(s); BSA: bovine serum albumin; C: cytidine; cDNA: DNA complementary to RNA; Ci: curie; dATP: 2'-deoxyadenosine-5'-triphosphate; dCTP: 2'-deoxy-cytidine-5'-triphosphate; ddATP: 2',3'-dideoxy-adenosine-5'-triphosphate; ddCTP: 2',3'-dideoxy-cytidine-5'-triphosphate; ddGTP: 2',3'-dideoxy-guanosine-5'-triphosphate; ddNTP: 2',3'-dideoxy-nucleoside-5'-triphosphate; ddTTP: 2',3'-dideoxy-thymidine-5'-triphosphate; dGTP: 2'-deoxy-guanosine-5'-triphosphate; dITP: 2'-deoxy-inosine-5'-triphosphate; DNA: deoxyribonucleic acid; dNTP: 2'-deoxy-nucleoside-5'-triphosphate; dTTP: 2'-deoxy-thymidine-5'-triphosphate; DTT: dithiothreitol; EDTA: (ethylenedinitrilo)tetraacetic acid; G: guanosine; g: gram/force of gravity; kb: kilobase(s); mg: milligram; ml: millilitre; mRNA messenger RNA; NAD: β-nicotinamide-adenine dinucleotide; NADP: β-nicotinamide-adenine dinucleotide phosphate; p: plasmid; RNA: ribonucleic acid; r.p.m: revolutions per minute; SDS: sodium dodecyl sulphate; T: thymidine; Tris: 2-amino-2-(hydroxymethyl)-1,3-propanediol; U: uridine; uCi: microcurie; ug: microgram; ul: microlitre; v/v: volume per volume; w/v: weight per volume.
LIST OF TABLES.

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.1 Genotypes of <em>Aspergillus</em> strains.</td>
<td>54</td>
</tr>
<tr>
<td>2.6.1 Plasmids used in this study.</td>
<td>55</td>
</tr>
<tr>
<td>2.7.1 dNTP solutions for DNA sequencing.</td>
<td>65</td>
</tr>
<tr>
<td>3.1.1 Spontaneous and induced &quot;reversion&quot; of creA&lt;sup&gt;d-1&lt;/sup&gt; and creA&lt;sup&gt;204&lt;/sup&gt;.</td>
<td>72</td>
</tr>
<tr>
<td>3.2.1 creA copy number in CreA&lt;sup&gt;+&lt;/sup&gt; transformants.</td>
<td>85</td>
</tr>
<tr>
<td>3.3.1 Relative levels of expression of creA in wildtype <em>A. nidulans</em> and strain 664.</td>
<td>92</td>
</tr>
<tr>
<td>3.3.2 Relative levels of expression of creA in creA mutant strains and CreA&lt;sup&gt;+&lt;/sup&gt; transformants.</td>
<td>92</td>
</tr>
<tr>
<td>4.1.1 Expected bands of hybridization of pANC4 to wildtype <em>A. nidulans</em> DNA.</td>
<td>100</td>
</tr>
<tr>
<td>5.1.1 Codon usage in the creA gene of <em>A. nidulans</em>.</td>
<td>127</td>
</tr>
</tbody>
</table>
LIST OF FIGURES.

FIGURE    PAGE

3.1.1    Spontaneous and induced "reversion" of cre\textsuperscript{d-1} and creA\textsubscript{204}. 70
3.1.2    Growth testing of T41a, T41b and T41c. 75
3.1.3    Complementation of C43 with pANCl. 77
3.1.4    Partial restriction map of pANCl. 78
3.1.5    Southern blot analysis of wildtype A. nidulans and T41. 80
3.1.6    Partial restriction map of the pANCl insert. 82
3.1.7    Southern blot analysis of wildtype A. nidulans DNA probed with pANCl. 83
3.1.8    Southern blot analysis of creA mutant alleles. 84
3.2.1    Southern blot analysis of A. nidulans pANCl transformants. 87
3.3.1    Northern blot analysis of wildtype A. nidulans and creA\textsubscript{204} probed with pANCl. 90
3.3.2    Northern blot analysis of wildtype A. nidulans probed with pANCl. 91
3.3.3    Northern blot analysis of creA mutant strains. 94
3.3.4    Northern blot analysis of CreA\textsuperscript{+} transformants. 96
4.1.1    Southern blot analysis of cre\textsuperscript{d-30}. 99
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.2</td>
<td>Northern blot analysis of creA&lt;sup&gt;d&lt;/sup&gt;-30.</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Partial restriction map of the insert in pANC6.</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Partial restriction map of the insert in pANC8.</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Northern blot analysis using pANC8 as a probe.</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Southern blot of pANC8 transformants of C62.</td>
</tr>
<tr>
<td>4.2.5</td>
<td>Strategy used for deleting the creA gene from A. nidulans.</td>
</tr>
<tr>
<td>4.2.6</td>
<td>Southern analyses of TC6330 and TC6332.</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Southern blot of DNA from various fungi probed with the creA clone.</td>
</tr>
<tr>
<td>5.1.1</td>
<td>creA sequencing strategy.</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Nucleotide sequence of creA.</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Primer extension analysis.</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Alignment of zinc finger domains.</td>
</tr>
<tr>
<td>5.2.2</td>
<td>The finger motifs proposed for zinc fingers I and II in CreA.</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Comparison of CreA and MIG1 zinc fingers.</td>
</tr>
<tr>
<td>5.2.4</td>
<td>Comparison of H-C links in zinc finger proteins.</td>
</tr>
<tr>
<td>5.2.5</td>
<td>Plotstructure of Peptidestructure for CreA.</td>
</tr>
<tr>
<td>5.2.6</td>
<td>Comparison of zinc finger loop regions in zinc finger proteins.</td>
</tr>
<tr>
<td>5.2.7</td>
<td>Comparison of CreA and Kruppel amino acid sequences.</td>
</tr>
<tr>
<td>5.2.8</td>
<td>Comparison of CreA and Engrailed amino acid sequences.</td>
</tr>
<tr>
<td>5.2.9</td>
<td>Comparison of CreA and Evenskipped amino acid sequences.</td>
</tr>
<tr>
<td>5.2.10</td>
<td>Comparison of CreA and Cyc8 amino acid sequences.</td>
</tr>
<tr>
<td>5.2.11</td>
<td>Chou and Fasman CreA secondary structure.</td>
</tr>
<tr>
<td>5.2.12</td>
<td>GOR CreA secondary structure.</td>
</tr>
</tbody>
</table>
CHAPTER ONE
INTRODUCTION.

Carbon catabolite repression, also referred to as glucose repression, is a wide domain regulatory system which allows the regulation of a large number of genes in response to the carbon source available. It functions to repress the expression of a large number of genes involved in the utilization of alternative carbon sources when a readily utilizable carbon source such as glucose or sucrose is also present, thus allowing more efficient use of resources. The metabolism of alternative carbon sources requires the induction and/or derepression of expression of subsets of genes for the enzymes required for a specific metabolic pathway, in response to particular environmental stimuli.

The mechanism of carbon catabolite repression is fairly well understood in *E. coli*. However, the molecular basis for this process in eukaryotes appears to be more complex and has not been well characterized. Considerable interest has been generated over the past few years for elucidating the basis for carbon catabolite repression at the molecular level. The study of carbon catabolite repression provides a means by which major systems of gene regulation may be understood from both a pure research perspective and also in the application of this knowledge to the control of gene expression in industrially important organisms. A large number of mutants showing an altered response to carbon catabolite repression have been
isolated in various simple eukaryotes. For example, at least a dozen putative regulatory genes involved in glucose repression in *Saccharomyces cerevisiae* have been identified by genetic analysis (reviewed by Entian 1986, Gancedo and Gancedo 1986). Although many of these have been genetically characterized, extensive studies at the molecular level are required to determine the sequence of events involved in the regulation of gene expression by carbon catabolite repression.

This thesis outlines a study of crea which is a wide domain regulatory gene involved in carbon catabolite repression in *Aspergillus nidulans*. This provides a model system for understanding some properties of major mechanisms of gene regulation in simple eukaryotes. There are many advantages to studying complex systems of gene regulation in *A. nidulans* as both classical genetic analyses and the molecular characterization of genes involved in specific regulatory pathways can be readily undertaken.

The filamentous ascomycete fungus, *A. nidulans*, is metabolically versatile and can grow on defined solid and liquid media. It has a haploid complement of eight chromosomes (Elliot 1960) and a large number of mutants have been mapped to the eight linkage groups (reviewed by Clutterbuck 1984) providing an extensive genetic map of this species. The existence of both sexual and parasexual stages in its life cycle allows both meiotic, mitotic and haploidization analyses. These are desirable features to complement molecular techniques for comprehensive studies into gene function and regulation. *A. nidulans* has a relatively small genome of about 3 X 10^7bp (Timberlake 1978, Brody and Carbon 1989), which facilitates the identification and isolation of particular sequences of interest.
from genomic libraries.

A number of transformation systems have been developed for *A. nidulans* and related species using selectable markers in either plasmid or cosmid vectors for homologous and non-homologous integration into the host genome. These include both the complementation of recessive auxotrophic mutations with cloned genes and the use of dominant selectable markers (reviewed by Gurr *et al.* 1987, Fincham 1989), allowing both the complementation analysis of cloned sequences and the analysis of mutations by reverse genetics. A further important feature of *Aspergillus* species is that cellular differentiation during sporulation provides a system to study the mechanisms by which a large number of genes are regulated both spatially and temporally (reviewed by Timberlake and Marshall 1988). Therefore, *A. nidulans* provides an amenable model for investigating the molecular mechanisms of wide domain regulatory systems in eukaryotes.

As a background to the study of the *creA* gene from *A. nidulans*, it is important to review some of the major mechanisms by which eukaryotic regulatory genes exert their effect at the molecular level. In addition some of the major observations and interpretations of the more widely studied genes involved in carbon catabolite repression in both prokaryotes and eukaryotes are reviewed.

1.1 MOLECULAR MECHANISMS OF GENE REGULATION IN EUKARYOTES:

The major control of gene regulation in eukaryotic cells is at the level of transcription (Darnell 1982, reviewed by Latchman 1990). This transcriptional regulation of eukaryotic genes is mediated by
the interaction of cis-acting DNA sequences with trans-acting factors (reviewed by Dynan and Tjian 1985). As many of the regulatory genes characterized from A. nidulans, including creA (this thesis) code for proteins having similar structural domains to known DNA binding proteins and have been shown to bind DNA, it is relevant to review the major characteristics of eukaryotic transcription factors.

1.1.1 DNA binding motifs:

Analyses of the structure and derived amino acid sequences of eukaryotic transcription factors have demonstrated that there are several motifs in these proteins which enable them to bind to DNA in a sequence specific manner (reviewed by Latchman 1990, reviewed by Nussinov 1990).

X-ray crystallography of the bacteriophage Cro protein (Anderson et al. 1981) the E. coli CRP (McKay and Steitz 1981) and the CI (Dodd and Egan 1987) proteins revealed the helix-turn-helix motif responsible for the ability of these proteins to bind in a sequence specific manner to DNA. This motif is present as the essential DNA binding domain in a number of regulatory proteins, including the proteins encoded by the homeotic genes of Drosophila melanogaster, referred to as homeodomains (Mihara and Kaiser 1988, Muller et al. 1988). Analysis of the 3-dimensional structure of the helix-turn-helix motif has shown that it consists of an α-helix followed by a turn followed by another α-helix in the protein. When interacting with DNA, the first α-helix lies across the major groove of the DNA target sequence while the second α-helix lies partly within the major groove and makes the sequence specific contact with
the DNA (Ptashne 1986).

Another motif is the leucine zipper (Landschulz et al. 1988a). This element has been shown to be present in a number of proteins including the mammalian enhancer binding protein C/EBP (Johnson et al. 1987), Gcn4 from yeast, the proto-oncogene proteins Myc, Fos and Jun and the Drosophila daughterless protein (Landschulz et al. 1988a, reviewed by Prendergast and Ziff 1989). The leucine zipper is characterized by a region containing four leucine residues positioned at intervals of seven amino acids in an α-helix, whereby the leucine residues occur every two turns on the same side of the helix. This structure is required for dimerization of the protein essential for subsequent DNA binding. The leucine zipper is therefore involved in protein-protein interactions and is not a DNA binding motif itself. Dimerization of the protein (or the formation of heterodimers) is mediated by the interlocking of two leucine containing helices on different molecules and results in the correct conformation for binding in the adjacent DNA binding region of the protein or protein complex (Johnson et al 1987, Landschulz et al. 1988a, b).

Zinc finger motifs were originally identified as DNA binding structures in the RNA polymerase III transcription factor TFIIIA which binds to the internal control region of the 5S RNA gene in Xenopus (Miller et al. 1985). Subsequently, at least two types of zinc fingers have been found in DNA binding factors involved in transcription mediated by RNA polymerase II. TFIIIA-like zinc fingers have been identified in a large number of protein sequences (reviewed by Berg 1990) including the mammalian transcription factor Sp1 (Kadonga et al. 1987), the Drosophila kruppel protein (Redmann et al. 1988) and the yeast AdrI protein (Parraga et al. 1988). The zinc
finger consensus sequence is Tyr/Phe-X-Cys-X₂₄-Cys-X₃-Phe-X₅-Leu-X₂-
His-X₃₄-His, where X is a variable amino acid (Miller et al. 1985). This sequence contains two invariant pairs of cysteine and histidine residues that stabilize the domain by tetrahedrally co-ordinating a Zn²⁺ ion. Protein modelling predicts that this would result in a finger-like structure where the conserved phenylalanine and leucine residues in the finger project from the surface of the protein. The tips of these fingers directly contact the major groove of DNA in the protein's target site where alternate fingers bind on opposite sides of the helix (Klug and Rhodes 1987). TFIIIA has nine tandemly arranged zinc finger motifs (Miller et al. 1985). The solution structure of a number of these zinc finger proteins have been identified using nuclear magnetic resonance data and confirm the major characteristics of the previously predicted models for these proteins involved in metal binding (Lee et al. 1989, Klevit et al. 1990, Omichinshi et al 1990).

A second class of zinc finger proteins are found in some steroid/thyroid hormone receptors which activate or repress gene expression by binding to specific DNA sequences in their target genes (Evans 1988, Beato 1989) and have also been observed in a number of other proteins such as the yeast transcription factor Gal4 (Giniger et al. 1985). The binding region of these fingers was originally thought to consist of four cysteine residues which co-ordinate the Zn²⁺ ion (Hollenberg et al. 1987). However, more recent studies of the metal binding domain of Gal4 which consists of a Cys-X₂-Cys-X₆-
Cys-X₆-Cys-X₂-Cys-X₆-Cys motif have suggested that instead of forming a zinc finger structure, the six cysteine residues form a binuclear metal ion cluster. Pan and Coleman (1990) have demonstrated that this
motif in Gal4 binds two metal ions which are coordinated by the six cysteines of the motif where two of these form connecting ligands between the ions. This alternative model for the structure of C2C2 zinc fingers is yet to be demonstrated for other transcription factors containing this motif, however the sequence conservation of such domains suggests that they are likely to form similar binuclear clusters.

1.1.2 Activation domains:

Following the binding of transcription factors to specific DNA sequences in the 5' region(s) of their target genes, these proteins must interact with other transcription factors or with RNA polymerase in order to influence transcription either in a positive way by activation (reviewed by Guarente 1988, reviewed by Ptashne and Gann 1990) or negatively by repression (reviewed by Levine and Manley 1989, reviewed by Renkawitz 1990).

In a number of cases specific acidic domains, distinct from those that mediate DNA binding, are required for transcriptional activation by positively acting transcription factors (Hope and Struhl 1986, Hollenberg and Evans 1988). These activation domains probably interact with other proteins such as the TATA box binding factor TFIID (Sawadogo and Roeder 1985) rather than by a direct interaction with RNA polymerase. For example, the binding of the yeast transcription factor Gal4 to its target site has been shown to alter the conformation of already bound TFIID, such that TFIID interacts with both the TATA box and the start point of transcription, rather than with the TATA box alone. The altered
binding of TFIID facilitates the binding of other transcription factors and RNA polymerase into a transcription complex and allows transcription to be initiated (Horikoshi et al. 1988).

1.1.3 Negative control systems:

Negatively acting transcription factors have been found to be less common in eukaryotic cells, however some such as the yeast Gal80 protein have been well characterized. Levine and Manley (1989) some of the mechanisms by which negative regulation may be achieved. The simplest mechanism of repression occurs where positively acting transcription factors are prevented from binding to their target sites by the binding of a repressor molecule. This is also common in bacteria, for example the lacI gene product blocks transcription of the Lac-operon (reviewed by Busby 1986). Similarly, a trans-acting protein may bind positive regulatory molecules preventing their binding to the transcription complex. A negatively acting factor may also interfere with the activation of transcription by an already bound transcription factor. For example, the negatively acting Gal80 protein from yeast prevents activation mediated by Gal4 of the structural genes for galactose utilization by binding to and masking the activation domain of Gal4 (Johnston et al. 1987). The interaction between Gal4 and Gal80 is discussed further in section 1.3.

Studies of eukaryotic gene regulation have focused mainly on positive regulatory sequences and positively acting transcription factors. However, a number of regulatory systems involving regulation by negatively acting factors (repressors) have been the subject of extensive analyses.
The homeodomains of *D. melanogaster* homeobox proteins mediate the binding of these proteins to specific DNA target sequences to regulate the transcription from target promoters. Most of these genes controlling early embryonic development in *Drosophila* function positively to activate transcription of their target genes. However, a number of these have also been shown to be transcriptional repressors, while others have repressor as well as activator function where the mode of activity appears to depend on the promoter they are regulating (reviewed by Hayashi and Scott 1990). Induction of the segment polarity gene *engrailed* by the homeobox proteins zerknullt-related, fushi tarazu and paired is inhibited by the *evenskipped* and *engrailed* gene products (Han *et al*. 1989) which recognize similar DNA target sequences (reviewed by Levine and Manley 1989). It has therefore been suggested that these proteins compete with positively acting factors for target sites to mediate their repressor function. Other studies have suggested that evenskipped may act as a repressor by directly interfering with the activities of general transcription factors, such as TFIID, which form the basal transcription complex (Ohkuma *et al*. 1990), rather than preventing an activator protein from binding to the same DNA sequence element. *In vitro* transcription of the *ultrabithorax* gene is similarly inhibited by evenskipped and it has been demonstrated that this repression is dependent on the interaction of evenskipped with specific sequences downstream of the *ultrabithorax* start point of transcription. Furthermore, these sites can also mediate repression by evenskipped when placed upstream of a heterologous promoter (Biggin and Tjian 1989). Evidence that repression by evenskipped is mediated by interference with the transcription complex, rather than by the competitive binding of transcription factors, comes from the finding that the binding of the
zerknult homeobox protein to these sites does not alter evenskipped's repression of ultrabithorax transcription (Biggin and Tjian 1989). Domains implicated in repressor activity by some of the D. melanogaster homeobox proteins are discussed further in Chapter 5.

In Neurospora crassa the expression of the structural genes involved in the utilization of nitrogen sources is regulated by the positively acting nit-2 gene (Fu and Marzluf 1987) as well as a negatively acting gene nmr (Premakumar et al. 1980, Dunn-Coleman et al. 1981). The nit-2 gene encodes a protein containing a zinc finger motif and is presumed to function as a DNA binding factor which activates transcription of the structural genes involved in nitrogen assimilation (Fu and Marzluf 1990). Mutations in the nmr gene lead to the constitutive expression of nitrogen related structural genes (Dunn-Coleman et al. 1981) and therefore probably acts as a repressor of the structural genes of the nitrogen circuit. nmr is expressed constitutively at low levels in N. crassa and does not appear to regulate nit-2 expression (Fu et al. 1988). nmr encodes a protein which has regions similar to the product of the AGRII gene from Saccharomyces cerevisiae, although unlike AgrII it does not contain DNA binding motifs, suggesting that Nmr is not a DNA binding protein which mediates transcriptional repression (Young et al. 1990, Jarai and Marzluf 1990). Alternative models for nmr's role in nitrogen repression propose that Nmr may bind to the Nit-2 protein which interferes with the activation of transcription by Nit-2, or the DNA binding domains of Nit-2 (Jarai and Marzluf 1990). The manner in which the Nit-2 and Nmr proteins interact remain to be elucidated.

The utilization of quinic acid as a carbon source in N. crassa is controlled by a cluster of five structural genes and two the
regulatory genes qa-IF and qa-IS (Giles et al. 1985). The genes of this cluster are induced by quinic acid and subject to glucose repression (Patel et al. 1981). Mutations in qa-IF lead to a non-inducable phenotype and are recessive (Case and Giles 1975). qa-IF has been shown to encode an activator which is positively required for transcription of all the qa genes including itself during induction (reviewed by Geever et al. 1989) which is consistent with the phenotype of qa-IF- mutants. The qa-IF protein activates transcription by binding to target sites which are present upstream of each qa gene in the cluster (Baum et al. 1987). Two classes of mutations in qa-IS have been identified and lead to either a dominant non-inducable phenotype (qa-IS-) or a recessive constitutive phenotype (qa-ISc) (Case and Giles 1975). These authors implied from the properties of the mutants that qa-IS codes for a repressor that interferes with the activation of transcription by Qa-1F and that the repressor itself is inhibited by quinic acid. The analyses of wildtype and mutant qa-IS nucleotide sequences are consistent with the hypothesis that qa-IS encodes a repressor. The protein encoded by qa-IS does not contain any motifs characteristic of DNA binding proteins (Huiet and Giles 1986) and therefore its target is more likely to be Qa-1F where it may abolish the ability of Qa-1F to bind to its target sequences or its ability to activate transcription at these promoters.

The qa cluster of N. crassa corresponds to the qut cluster which encodes six genes involved in quinic acid utilization in A. nidulans (reviewed by Grant et al. 1988). Regulation of the qut cluster is mediated similarly by a positively acting regulatory protein within the cluster, qutA, and a negatively acting regulator encoded by the
closely linked qutR gene (Beri et al. 1987, Grant et al. 1988). These are the A. nidulans equivalents of qa-1F and qa-1S respectively. Like qa-1F, qutA is presumed to activate transcription of the qut structural genes by binding to target sites upstream of these genes. Potential QutA target sites have been identified and await further analyses (Hawkins et al. 1988). Mutations in the qutR locus result in induced levels of expression of the structural genes, but expression remains subject to carbon catabolite repression in qutR- strains. The mechanisms by which Qa-1S and QutR act as repressor at the molecular level remain to be determined and their targets identified.

1.1.4 The regulation of synthesis and activity of transcription factors:

In order for specific gene regulation mediated by transcription factors to occur, the factors themselves must respond to specific environmental stimuli. A number of mechanisms exist which may regulate the specific activities of transcription factors. The expression of genes encoding transcription factors may be regulated themselves so that the synthesis of the encoded proteins only occurs under conditions for which they are required or in specific cell types. This is very common, for example, in cell differentiation in higher eukaryotes (reviewed by Mitchell and Tjian 1989). The synthesis of eukaryotic transcription factors is often regulated post-transcriptionally. For example, the synthesis of Gcn4 which activates genes encoding enzymes for the biosynthesis of amino acids in yeast cells occurs in response to amino acid starvation. Synthesis of Gcn4 is mediated via increased translation of its mRNA. Regulation is made possible by the inhibition of translation initiation at AUG
codons upstream from the correct initiation site, leading to an increase in translation from the correct start codon in response to amino acid starvation (Fink 1986).

Other transcription factors may be expressed or synthesized constitutively in cells and may be activated/inactivated at the post-translational level by ligand binding (for example, the modification of the yeast AceI protein by copper – Furst et al. 1988), protein modifications via glycosylation (Jackson and Tjian 1988) or phosphorylation (Sorger and Pelham 1988) or by a mechanism which disrupts protein-protein interactions such as that seen for the Gal80 and Gal4 regulatory proteins in response to galactose (see section 1.3.1).

1.2 CARBON CATABOLITE REPRESSION IN Escherichia coli:

The most extensively studied example of carbon catabolite repression in prokaryotic organisms is that of the regulation of the lacZ operon in Escherichia coli. A great deal is now known about how the regulatory molecules involved in the utilization of lactose as a carbon source act at the molecular level to activate and repress the expression of the lactose operon (reviewed by Busby 1986).

The lactose operon is inducible, and subject to carbon catabolite repression. The operon is negatively regulated by the product of the lacI gene, the Lac repressor protein. In the absence of an inducer the Lac repressor protein binds to one major and two auxiliary operator sites in the lac promoter (Reznikoff et al. 1974, Gilbert et al. 1976, Fried and Crothers 1981) which hinders the binding of RNA polymerase and therefore the initiation of
transcription, resulting in repression of the lac operon. Binding to one of the minor operator sites may also prevent the binding of the cyclic-3',5'-adenosine monophosphate (cAMP) receptor protein (CRP), which is positively required for the initiation of transcription of the lactose operon (Oehler et al. 1990). However, when Lac repressor protein is complexed with an inducer, such as lactose or allolactose, it can no longer bind to the major operator site to prevent transcription and derepressed levels of transcription occur.

The expression of the lactose operon is also subject to positive regulation via CRP, which also regulates the expression of other operons which are glucose sensitive and involved in the utilization of alternative carbon sources. Regulation by carbon catabolite repression in *E. coli* acts at the level of transcription, resulting in decreased synthesis of lac mRNA (Varmus 1970a and 1970b). The operons which are subject to regulation by CRP have been collectively named the cAMP - CRP regulon (Magasanik and Neidhardt 1987). When *E. coli* cells are grown in the presence of lactose as the carbon source the intracellular concentration of cAMP rises rapidly and upon the addition of glucose to the medium this concentration decreases immediately (reviewed by Ullmann and Danchin 1983). If the level of cAMP within cells is high, cAMP binds to CRP. This cAMP - CRP complex assumes a new conformation, which allows it to bind to specific DNA sequences in or near catabolite sensitive promoters including the promoter region of the lactose operon (reviewed by de Crombegghe et al. 1984). Such binding allows the initiation of transcription of the lactose operon. During growth of *E. coli* in medium containing glucose the intracellular level of cAMP drops and cAMP - CRP mediated transcription is suppressed. The CRP molecule is composed of two
identical subunits (Anderson et al. 1971, Aiba et al. 1982, Cossart and Gicquel-Sanzey 1982) enabling two molecules of cAMP to bind to this dimeric protein, one molecule per subunit (Weber et al. 1982). Each CRP subunit is composed of two structural domains. The COOH terminal domain contains a DNA binding helix-turn-helix motif and the NH$_2$-terminal domain contains a single cAMP binding site and a subunit interaction site (Krakow and Pastan 1973, Eilen et al. 1978, Aiba and Krakow 1981, McKay and Steitz 1981). Chemical protection experiments with the lac promoter indicate that the protein has its major contacts in two successive major grooves in the target DNA binding site (Simpson 1980). The amino acid residues of CRP which are directly involved in making sequence specific contacts with this DNA have been identified (Ebright et al. 1984). Comparisons of a number of cAMP - CRP binding sites has shown that these sites span 22bp with the consensus sequence being 5' aaaaTGTGATctagATCACATtt 3', containing an inverted repeat of an 11bp element where the upper case letters represent the highly conserved nucleotides (Jayaraman et al. 1989). The distance between the start point of transcription and the CRP binding site within CRP responsive promoters has been shown not to be critical and therefore there is probably not direct contact between the cAMP - CRP complex and the RNA polymerase molecule in order for the initiation of transcription (Ullmann and Danchin 1983). When cAMP - CRP binds to its target site, the DNA sequences flanking this site wrap around the CRP which stabilizes the binding (Unger. et al 1983) and probably causes the DNA to bend in the *E. coli* lac promoter (Wu and Crothers 1984). The wildtype lac promoter (P1) is an inefficient promoter in the absence of CRP and cAMP. A second lac promoter P2 has been identified in the lac operon regulatory region which leads to the initiation of transcription at low frequency 22bp
upstream from the major point of transcription (Hawley et al. 1982, Reznikoff et al. 1982). It has therefore been suggested that RNA polymerase I has a greater affinity for P2 and other promoters in the absence of cAMP-CRP. However, in the presence of cAMP-CRP, the binding of this complex to P1 inhibits transcription from the less efficient promoter P2 by blocking RNA polymerase binding. Under these conditions P1 no longer competes for RNA polymerase binding and transcription is initiated at high frequency from the major start point (Hawley et al. 1982). Therefore, cAMP-CRP activates Lac transcription by both increasing the binding rate of RNA polymerase and its affinity to P1, and by excluding competing RNA polymerase binding sites.

Adenylate cyclase has been shown to be the sensor protein which is activated/inactivated in response to whether the carbon source is glucose or a poor one and converts ATP to cAMP accordingly. Adenylate cyclase is activated/inactivated by a process involving phosphorylation-dephosphorylation of the sugar phosphotransferase system, although the precise mechanism is not known (Roseman and Meadow 1990).

1.3 CARBON CATABOLITE REPRESSION IN Saccharomyces:

Studies of carbon catabolite repression in simple eukaryotes have mainly concentrated on the regulation of carbohydrate metabolism in the yeast Saccharomyces cerevisiae. In the presence of glucose the synthesis of many enzymes is repressed, particularly those involved in gluconeogenesis, the tricarboxylic acid cycle, the glyoxylate cycle and the catabolism of exogenously supplied sugars such as
maltose, sucrose and galactose (reviewed by Wills 1990). There is a growing amount of genetical and biochemical evidence that the role of cAMP in carbon catabolite repression in *S. cerevisiae* does not have the same direct role as that found in *E. coli* (Matsumoto *et al.* 1983a, Eraso and Gancedo 1984). It has been shown that intracellular levels of cAMP increase transiently in response to glucose in the growth medium. This pulse of cAMP may be sufficient to start a cascade effect, bringing about the activation/inactivation of cAMP dependent protein kinases, which in turn leads to either repression or synthesis of enzymes involved in carbon metabolism (Beullens *et al.* 1988). *A priori*, there is no reason to believe that mechanisms of carbon catabolite repression are conserved between prokaryotes and eukaryotes.

Different strains of *S. cerevisiae* are extremely specialized in their ability to metabolize both fermentable and non-fermentable substrates. This specialization has resulted from the strong selection for industrial strains capable of metabolizing specific carbon sources efficiently for the production of various metabolites such as ethanol. In *S. cerevisiae*, families of genes are involved in the utilization of various carbon sources. Therefore, different strains may have fixed allelic differences and gene family loci may vary between strains (reviewed by Carlson 1986). Not surprisingly, there are many conflicting reports on the ability of various strains to utilize specific carbon sources and on the specific regulatory mechanisms involved in the metabolism of such carbon sources.

A large number of mutations have been identified in *S. cerevisiae* which lead to an altered response to carbon catabolite repression. Some of these are allelic despite showing quite different
phenotypic effects (reviewed by Gancedo and Gancedo 1986). With studies now focusing on the study of these genes at the molecular level, the complexities of the regulation of structural genes by a number of regulatory systems are beginning to emerge. The regulatory systems can be positive or negative (reviewed by Guarente 1984) and encompass a number of specific and global regulatory molecules effecting a response in a hierarchical manner. The study of carbon catabolite repression in yeast is further complicated by the finding that the products of some genes are required for normal repression in the presence of glucose while the products of other genes are required for the derepression process (reviewed by Gancedo and Gancedo 1986). Another complicating factor is that while the synthesis of some enzymes are subject to carbon catabolite repression, these enzymes may also be readily inactivated by the addition of glucose, that is, subject to catabolite inactivation (reviewed by Holzer 1976).

1.3.1 Regulation of the galactose regulon of *S. cerevisiae*:

The regulation of the structural genes involved in the utilization of galactose as a carbon source has been extensively studied in *S. cerevisiae* and provides a good example of control by multiple regulatory factors.

The structural genes *GAL1*, *GAL7* and *GAL10*, which are required for the utilization of galactose, are subject to galactose regulon specific induction by both positively and negatively acting regulatory molecules (reviewed by Johnston 1987). Positive regulation is effected by the product of the *GAL4* gene. The Gal4 protein is a
DNA binding protein which binds in a cooperative manner to four upstream activating sequences (UAS) of 17bp in length, upstream from the structural genes (Bram and Kornberg 1985, Giniger et al. 1985, Bram et al. 1986, Tajima et al. 1986) and activates transcription from these promoters. Gal80 is a negative regulator of gal-regulon transcription. GAL80 transcription is subject to induction by the presence of galactose although the mechanism is not understood. When cells are grown in the absence of inducer the Gal80 protein complexes with Gal4, preventing it from activating transcription of the structural genes. However in the presence of galactose, either galactose itself or a metabolite of galactose, causes this complex to disassociate freeing Gal4 which can then activate transcription (reviewed by Oshima 1982, reviewed by Wills 1990). The association between Gal4 and Gal80 probably occurs while Gal4 is bound to its target sites, since it has been found to be bound to UAS's in both induced and uninduced growth conditions (Giniger et al. 1985, Lohr and Hopper 1985).

Expression of the GAL genes required for the metabolism of galactose in S. cerevisiae is stringently regulated by the carbon source available to the cells (reviewed by Johnston 1987). When glucose is present Gal4 disassociates from its UAS's and the galactose regulon is repressed (Matsumoto et al. 1983, Selleck and Majors 1987). However, the mechanism by which growth on glucose causes repression of the GAL and other genes is not well understood. A number of genes have been identified by mutational analysis to be involved in carbon catabolite repression of the galactose regulon. Mutations in HEX2, encoding hexokinase PII which catalyzes the phosphorylation of glucose, lead to the insensitivity of the GAL, SUC
(sucrose utilization) and MAL (maltose utilization) gene families to glucose repression (Zimmermann and Scheel 1977, Entian 1980, Ma and Botstein 1986). Such pleiotropic effects suggest that HEX2 has a global regulatory role in glucose repression. Mutations in the genes REG1 and GRR1 also lead to the loss of carbon catabolite repression of the galactose regulon (Matsumoto et al. 1983, Bailey and Woodword 1984, Neigeborn and Carlson 1987). GAL82 and GAL83 have an undefined role in mediating repression and mutations in these genes lead to the loss of carbon catabolite repression of the GAL genes specifically (Matsumoto et al. 1981, Matsumoto et al. 1983). Along with the genes required for normal repression, six SNF genes have been found to be required for the derepression of genes subject to carbon catabolite repression (Carlson et al. 1981, Neigeborn and Carlson 1984, Celenza and Carlson 1986). SNF1 encodes a protein kinase, further suggesting that protein phosphorylation has a major role in carbon catabolite repression in S. cerevisiae (Celenza and Carlson 1986).

Molecular analysis of the GAL1 promoter region has identified two elements which are involved in carbon catabolite repression of GAL1 transcription (Flick and Johnston 1990). The first element UASG, inhibits the activation of transcription of GAL1 by Gal4. The second element URS, of which there are probably multiple sequences, mediates glucose repression of GAL1 transcription which is independent of Gal4. GAL83, REG1, GRR1, SSN6 and SNF1 are required for repression at both of these elements, while GAL82 and HXX2 are required for the glucose repression independent of GAL4. Therefore, glucose repression of GAL1 appears to be conferred by two independent pathways which respond to growth on glucose (Flick and Johnston 1990). A third regulatory circuit may involve the competitive interaction between
the effectors glucose and galactose on the Gal80 protein in the cytoplasm or by a mechanism which lowers the intracellular concentration of gal-regulon inducer via the inhibition of gal-permease (Matern and Holzer 1977). These regulatory circuits have been incorporated into a model for carbon catabolite repression of the galactose regulon by Matsumoto et al. (1983), but the details of the mechanisms involved remain unclear.

The Lac9 protein, which positively regulates the galactose-lactose regulon of Kluyveromyces lactis, can partially substitute for Gal4 function in S. cerevisiae (Wray et al. 1987) and vice versa (Riley et al. 1987). The proteins have three homologous domains, one of which is the DNA binding domain, and have similar target sequences (Wray et al. 1987). As discussed above, the gal-regulon of S. cerevisiae is subject to glucose repression, however the gal-lac-regulon of K. lactis is not subject to such severe glucose repression (Dickson and Markin 1980). Interestingly, complementation of lac9 with GAL4 leads to a glucose repressible gal-lac regulon in K. lactis (Riley et al. 1987) and complementation of a gal4 defective strain with Lac9 leads to glucose repression of the gal-regulon in S. cerevisiae which is less severe than in strains containing GAL4 (Wray et al. 1987). This suggests that Lac9 function may have some response to the presence of glucose not previously detected. It has recently been confirmed that glucose repression is mediated by Lac9 in strains having Lac genes subject to glucose repression and that this activity is separate from its ability to activate transcription of the gal-lac regulon of K. lactis (Breunig 1989). Genetic evidence suggested that an analogous gene to the S. cerevisiae GAL80 gene functions to negatively regulate the gal-lac-regulon of K. lactis (designated
LAC10) (Dickson et al. 1990). It has subsequently been demonstrated that Lac9 can interact with Gal80 (Salmeron et al. 1989). Molecular and mutational analyses of a glucose sensitive LAC9 allele show that Lac9 is a limiting factor for β-galactosidase gene expression and that a negative regulatory protein like Gal80, may interact with Lac9 in the presence of glucose, since moderate overproduction of this Lac9 protein was found to be sufficient to overcome inhibitory effects of glucose (Kuger et al. 1990).

1.3.2 Sucrose utilization in S. cerevisiae:

The SUC2 gene from S. cerevisiae encodes a secreted invertase which is responsible for the extracellular hydrolysis of sucrose (Carlson and Botstein 1982). The SUC2 gene is particularly useful for studying carbon catabolite repression since this is the only regulatory mechanism controlling SUC2 expression. SUC2 expression is not subject to regulation by induction by the substrate as is the case with most of the structural genes involved in the utilization of alternative carbon sources (Carlson and Botstein 1982). However, the regulation of SUC2 by carbon catabolite repression remains a complex system. Carlson (1987) has reviewed the genes which affect glucose repression of SUC2, of which at least twelve have been identified. These include SNF1-5 and SNF7, REG1, CID1, SSN6, TUP1, (reviewed by Carlson 1987) and RGR1 (Sakai et al. 1990). Mutations in all of these genes can also lead to altered regulation of other genes subject to carbon catabolite repression.

Analysis of the SUC2 promoter has identified a region which confers glucose-repressible expression of the SUC2 gene (Sarokin and...
Carlson 1984) and mutations in the trans-acting genes SNF1-5 and SNF7 partially or completely block derepression of SUC2 gene expression (Carlson et al. 1981, Neighborn and Carlson 1984). Further analysis of SNF2 and SNF5 at the molecular level using gene disruptions suggested that these genes are required for high level expression of SUC2, but are not directly involved in regulation (Abrams et al. 1986). Although SNF5 was identified as a gene required for Suc2 expression (Neigeborn and Carlson 1984), further studies suggest that SNF5 has a more general role in transcription rather than being restricted to glucose repression. For example, SNF5 is also required for derepression of acid phosphatase during phosphate starvation (Abrams et al. 1986) and mutations in SNF5 also lead to increase expression of protease B in stationary phase cells (Moehle and Jones 1990). A role for SNF5 in transcription is also supported by the finding that mutations in the essential gene SPT6 affect transcription and are able to restore regulation of invertase expression in snf5 mutations (Neigeborn et al. 1986, reviewed by Laurent et al. 1990). SNF5 has been cloned and sequenced (Laurent et al. 1990) and found to encode a glutamine and proline rich protein which has trans-activator function without the capacity to bind to DNA itself. Other associated proteins such as Snf2 and Snf6 may be required for the activation function of Snf5. Snf2, Snf5 and Snf6 are proposed by Laurent et al. (1990) to form a heteromeric complex where Snf5 provides a transcriptional activation domain and Snf2 and Snf6 may provide DNA binding activity. No evidence has yet been reported that Snf2 or Snf6 are capable of binding upstream of SUC2.

The SNF3 gene encodes a high affinity glucose transporter homologous to the mammalian protein (Celenza et al. 1988) and exerts
its effect early in the regulatory circuit. Interestingly, disruption of SNF3 reduces growth on sucrose and raffinose but has no effect on SUC2 expression (Neigeborn et al. 1986b), while SNF3 missense mutations lead to altered regulation of SUC2 expression (Neigeborn and Carlson 1984). The role of SNF3 in carbon catabolite repression is further supported by the fact that there is some evidence that high affinity glucose transport in S. cerevisiae is associated with hexokinase PII, which is probably one of the regulatory proteins involved in glucose repression (Entian and Frolich 1984). Genetic studies suggest that SNF1 (which encodes a serine/threonine protein kinase) and SNF4 are functionally related genes. Like SNF1, SNF4 is also required for the expression of many glucose repressible genes. Both SNF1 and SNF4 act at the transcriptional level of SUC2 expression (Carlson and Botstein 1982). More recently Celenza et al. (1989) have provided further genetic and biochemical evidence which suggest that Snf4 is a positive effector of the protein kinase encoded by SNF1 and that the two proteins are probably physically associated. Celenza et al. (1989) propose that either Snf4 stimulates Snf1 via direct protein-protein interactions or that Snf4 inhibits a negative effector of Snf1.

Mutations in SSN6 and TUP1 can suppress the derepressed phenotype of SNF1 mutant strains. It has been suggested, from studying the epistatic relationships between different classes of mutants, that SSN6 and TUP1 have direct negative regulatory effects on carbon catabolite repression. ssn6 mutations lead to high level, glucose insensitive expression of SUC2 and other glucose repressible genes (Schultz and Carlson 1987). SSN6 encodes a nuclear phosphoprotein which contains long tracts of poly(glutamine/alanine)
residues and by analogy to other proteins containing these tracts, Ssn6 is thought to be involved in transcriptional repression. The protein encoded by TUP1 also contains polyglutamine tracts. No DNA binding activity has been identified for Ssn6 or Tup1 and therefore it is possible that these proteins associate with or affect the activity of DNA binding proteins, leading to transcriptional repression.

A further gene, MIG1, is involved in glucose repression of SUC2 (Nehlin and Ronne 1990). Overexpression of MIG1 inhibits SUC2 expression whereas a MIG1 gene disruption leads to altered glucose repression of SUC2. MIG1 encodes a protein containing a C$_2$H$_2$ zinc-finger motif and has been found to bind to specific target sites upstream of SUC2. Mig1 is therefore believed to be a DNA binding trans-repressor of SUC2, however the molecular analysis of MIG1 suggests that it may also have an activator function which has not yet been identified (Nehlin and Ronne 1990).

It has been proposed that other genes such as CID1, REG1 and HXK2 are probably involved in the sensory processes of carbon catabolite repression and lead to an appropriate response by the cell to the carbon source available for utilization. Mutations in RGR1 lead, to the expression of SUC2 resistant to glucose repression, and to increased expression under glucose derepressing conditions (Sakai et al. 1988). Molecular analysis of RGR1 has shown that it is an essential gene and affects SUC2 expression possibly via an interaction with the pathway which regulates intercellular levels of cAMP (Sakai et al. 1990).

The regulation of the genes encoding enzymes necessary for the
utilization of various carbon sources in *S. cerevisiae* is via a complex series of regulatory genes which affect the induction and/or their regulation by carbon catabolite repression. Although many genes involved in carbon catabolite repression have been identified by mutation, little is known of the proteins they code for and how these molecules interact and function at the molecular level.

**1.4 GLOBAL GENE REGULATION IN A. nidulans:**

The most widely studied and best characterized mechanism of global gene regulation in *A. nidulans* and *N. crassa* is that of nitrogen metabolite repression. In these fungi the synthesis of a large number of enzymes involved in the utilization of alternative nitrogen sources are repressed by ammonium and glutamine (Arst and Cove 1973, reviewed by Marzluf 1981, reviewed by Wiame et al. 1985).

The positively acting regulatory gene *areA* mediates nitrogen metabolite repression in *A. nidulans*. The *areA* gene product is thought to act at the level of transcription and is required for the expression of structural genes subject to nitrogen metabolite repression. In wildtype *A. nidulans* glutamine and ammonium prevent the expression and/or activity of the *areA* gene product (reviewed by Arst and Scaccocchio 1985, reviewed by Wiame et al. 1985). A large number of *areA* mutant alleles, with diverse phenotypic effects, have been identified and characterized. Loss of function mutations are the most common alleles and lead to an inability to utilize nitrogen sources other than ammonium or glutamine and to repressed levels of enzymes subject to control by *areA* (reviewed by Arst and Cove 1973, reviewed by Wiame et al. 1985). Some of these mutants have been found
to be the result of translocations which interrupt the reading frame while another was found to be a chain termination mutant (Kudla et al. 1990). Other rarer alleles, such as the areA-102 mutation, result in elevated levels of some enzymes subject to nitrogen metabolite repression (Hynes and Pateman 1970, Hynes 1975a, Arst 1977) in addition to repressed levels for other enzyme activities regardless of the presence of ammonium or glutamine (Arst and Cove 1973, Arst and Scazzocchio 1975, Hynes 1975a). This phenotype suggests that receptor sites for the areA gene product may differ in structure and therefore affinity for the areA protein, allowing it to activate areA regulated genes with different efficiencies (reviewed by Wiane et al. 1985). areA has been cloned and sequenced (Caddick et al. 1986, Kudla et al. 1990) and the derived amino acid sequence contains a putative DNA binding zinc-finger of the Cys-X2-Cys-X17-Cys-X2-Cys type. The sequence is rich in the S(T)PXX motif which is characteristic for DNA binding proteins (Suzuki 1989). areA also has a highly acidic region which may be similar in structure to regions of regulatory proteins having a transcriptional activation function (Giniger and Ptashne 1987, Hope et al. 1988, Kudla et al. 1990). Sequence analysis of a number of areA mutant alleles has identified the regions of the putative protein that are essential for areA function and specificity. The areA-102 mutation was found to be a single amino acid change in a conserved residue of the areA zinc finger, demonstrating that particular residues within the loop structure are probably involved in the specific recognition of areA target sites and its affinity to these sites (Kudla et al. 1990).

The corresponding gene to areA in N. crassa is nit-2 (Reinert and Marzluf 1975, Coddington 1976, Dunn-Coleman et al. 1979). Like
areA, nit-2 is one of the major positively acting regulatory genes in *N. crassa* which regulates the expression of structural genes in response to the nitrogen source available for utilization (Marzluf 1981, Fu and Marzluf 1987). nit-2 was characterized by loss of function mutations and although derepressed alleles have not been isolated (Marzluf 1981), there seem to be functional similarities from the study of mutants between nit-2 and areA. The nit-2 gene complements areA- mutants of *A. nidulans* and turn on the expression of nitrate reductase, acetamidase and other enzymes in *A. nidulans* (Davis and Hynes 1987). This demonstrates that the activation function and DNA specificity of the nit-2 and areA gene products are quite similar and that the upstream recognition elements of the genes involved in nitrogen assimilation must be similar in these two fungi. The nit-2 gene has been sequenced and encodes a protein predicted to contain a single zinc finger which is similar to that predicted for the areA protein (Fu and Marzluf 1990a). Furthermore, gel retardation experiments and DNA footprinting studies have demonstrated that the Nit2 protein binds in a sequence specific manner to at least three sites in the 5' region of the *N. crassa* nitrate reductase gene and to related sites in the upstream promoter regions of the nitrate and nitrite reductase genes of *A. nidulans* (Fu and Marzluf 1990b). However, the mechanism of nitrogen metabolite repression in these fungi may not be completely alike.

nmr has been identified to be another major gene involved in nitrogen metabolite repression in *N. crassa* which has a negative function and is expressed constitutively. In nmr mutants a number of enzymes involved in nitrogen assimilation are expressed constitutively (Premakumar *et al*. 1980, Dunn-Coleman *et al*. 1981).
nnr has been cloned (Fu et al. 1988) and sequence analysis of nnr has not revealed any DNA binding motifs in the putative protein. However, the predicted Nmr protein shows sequence similarity to the negatively acting regulatory protein ArgrII involved in arginine synthesis repression in yeast (Young et al. 1990). It has been suggested that while *A. nidulans* does not appear to have a gene equivalent to *nnr*, *areA* may fulfill the functions of both the *N. crassa* genes *nit-2* and *nnr*. This is supported by the fact that some *areA* alleles have constitutive expression of nitrogen assimilation enzymes while no *nit-2* mutants lead to derepression and *nnr* mutant alleles show the derepressed phenotype (Fu et al. 1988).

Other wide domain regulatory genes have been identified in *A. nidulans*. These include *suAmeth* and *palcA* which appear to be negatively acting regulatory genes involved in sulphur repression, and the regulation of enzymes involved in the catabolism of phosphate sources respectively. Other genes, such as *pacC*, *palA*, *palB*, *palC*, *palE* and *palF* have been shown to be involved in the regulation of cellular activities in response to changes in pH (Arst and Scazzocchio 1985, Caddick et al. 1986a, 1986b). These global mechanisms of gene regulation in *A. nidulans* are less well characterized and the roles that the products of these genes play remain obscure. Although more is known about nitrogen metabolite and carbon catabolite repression (discussed in the following section) in *A. nidulans* there is limited understanding of how the regulatory genes or their products themselves are activated and inactivated in response to the carbon and nitrogen sources available for utilization. Furthermore, the expression of some genes involved in the utilization of compounds as both nitrogen and carbon sources such
as acetamide are subject to both nitrogen metabolite repression and carbon catabolite repression and relief of either of these two repression mechanisms leads to derepressed expression of \( \textit{amds} \) (Hynes 1970). The mechanisms by which these two forms of repression interact with each other at the molecular level are yet to be understood.

1.5 CARBON CATABOLITE REPRESSION IN \textit{A. nidulans}:

When \textit{A. nidulans} is grown in the presence of a source of carbon catabolite repression such as glucose or sucrose, the synthesis of a range of enzymes involved in the utilization of alternative carbon sources is decreased. Reduced activity has been demonstrated for a large number of enzymes under repressing conditions including: alcohol dehydrogenase (Page and Cove 1972), acetamidase (Hynes 1970, Hynes and Pateman 1970a and 1970b), proline permease and proline oxidase (Arst and MacDonald 1975, Bailey and Arst 1975), NAD-linked glutamate dehydrogenase (Kinghorn and Pateman 1973, Bailey and Arst 1975, Hynes 1974), intracellular and extracellular proteases (Cohen 1973), D-quinate dehydrogenase, \( \beta \)-galactosidase, \( \alpha \)-glucosidase, acetyl-CoA synthetase and isocitrate lyase (Hynes and Kelly 1977, Kelly and Hynes 1977). Further to this, it has been shown in some systems that carbon catabolite repression in \textit{A. nidulans} acts at the level of mRNA probably by controlling transcription, in for example, \textit{pelA} (Dean and Timberlake 1989), \textit{facB} (Katz and Hynes 1989a), \textit{alcA} and \textit{aldA} (Felenbok et al. 1989), \textit{alcR} (Lockington et al. 1987) and \textit{lmaA} and \textit{lmb} (Katz and Hynes 1989b).

Catabolite inactivation is the process by which there is rapid inactivation of enzyme activity as opposed to repression of
synthesis, after the addition of D-glucose or sucrose to the growth medium. Carbon catabolite repression and catabolite inactivation occur independently and by different mechanisms in yeast cells (reviewed by Holzer 1976). This independence was demonstrated by the characterization of yeast strains in which catabolite inactivation is defective while the response to carbon catabolite repression remained unaltered (Entian 1977). The presence of these two independent processes complicates the interpretation of the mutant phenotypes showing altered responses to carbon metabolism and assimilation in yeast. It has been shown that the activities of phosphoenolpyruvate carboxylase, fructose-1,6-diphosphatase, isocitrate lyase, acetyl CoA synthetase, malate synthase, NADP-malic enzyme, isocitrate dehydrogenase, malate dehydrogenase, alcohol dehydrogenase and acetamidase are not altered by D-glucose or sucrose in A. nidulans (Kelly 1980). The absence of catabolite inactivation (at least for these enzymes) has therefore allowed for the easier interpretation of mutants affected in the regulation of carbon catabolism in A. nidulans.

As previously outlined, the role of cAMP in carbon catabolite repression in yeast remains unclear. This is also the case with A. nidulans. The exogenous addition of adenylate cyclase activators such as glucagon and isoproterenol or inhibitors of cAMP phosphodiesterases such theophylline and aminophenline do not alter the levels at which enzymes subject to carbon catabolite repression are synthesized (Arst and Bailey 1977).
1.5.1 Selection of mutations affecting carbon catabolite repression in A. nidulans:

A number of methods have been used to isolate strains of A. nidulans showing derepressed levels of various enzymes in the presence of D-glucose and are therefore no longer subject to carbon catabolite repression.

Loss of function mutations in the areA, gene which codes for a positively acting regulatory protein involved in ammonium repression (Arst and Cove 1973), lead to an inability to utilize sole nitrogen sources other than ammonium (discussed in section 1.4). These mutants fail to derepress enzymes required for the utilization of other nitrogen sources even in the absence of a source of nitrogen repression. Enzymes for the utilization of compounds which can be used as both nitrogen and carbon sources, such as acetamide and L-proline are subject to both carbon catabolite repression and nitrogen metabolite repression. The relief of either of these two mechanisms of repression enables enzyme synthesis to occur (Hynes 1970, Arst and MacDonald 1975). Therefore, areA− mutant strains are able to utilize acetamide and L-proline as sole carbon and nitrogen sources (Arst and Cove 1973) but not in the presence of D-glucose or sucrose, due to carbon catabolite repression of the synthesis of the enzymes required to metabolize these compounds. Selection for growth of areA− mutant strains on media containing D-glucose with acetamide or L-proline as the nitrogen sources has been used to isolate A. nidulans strains having mutations resulting in the synthesis of acetamidase and enzymes for L-proline metabolism which are no longer sensitive to carbon catabolite repression by glucose (Arst and Cove 1973, Bailey
Another selection method utilizes strains which are mutant for pyruvate dehydrogenase activity. Pyruvate dehydrogenase mutants are unable to produce acetyl Co.A from pyruvate and therefore require an alternative source of acetyl CoA (Romano and Kornberg 1968, 1969). This requirement can be supplemented by the addition and metabolism of acetamide or ethanol as the sole carbon source in the absence of D-glucose or in the presence of derepressing carbon sources, such as glycerol, lactose, melibiose or L-arabinose. The metabolism of acetamide or ethanol can only provide a source of acetyl CoA in the absence of D-glucose, since acetamidase and alcohol dehydrogenase synthesis are regulated by carbon catabolite repression (Bailey and Arst 1975). Therefore, medium containing D-glucose together with acetamide or ethanol has been used to select for pdhA phenotypically revertant strains in which ethanol or acetamide could be utilized in the presence of D-glucose, thereby identifying strains in which alcohol dehydrogenase and acetamidase synthesis are no longer sensitive to repression by D-glucose (Bailey and Arst 1975, Arst and Bailey 1977).

Other mutations have been identified in strains already affected in carbon catabolite repression by selecting for the ability of further mutations to suppress the phenotype of the parent strain. The presence of derepressed levels of the enzymes alcohol dehydrogenase, acetyl CoA synthetase and acetamidase in creA mutants can be scored by their sensitivity to allyl alcohol, fluroacetate and fluroacetamide in the presence of D-glucose or sucrose. The phenotypes shown by creA alleles on these media abolishes the need to score derepression of these enzymes in areA:creA and pdhA:creA double
mutants, or assaying for enzyme activity. Allyl alcohol is toxic to \textit{A. nidulans} when converted, via alcohol dehydrogenase to its toxic product acrolein. \textit{creA} mutations lead to complete sensitivity to allyl alcohol in the presence of D-glucose or sucrose, while wildtype strains are resistant due to the carbon catabolite repression of alcohol dehydrogenase gene expression (Hynes and Kelly 1977). While wildtype strains are sensitive, \textit{creA} mutations lead to hypersensitivity to fluoroacetate and fluoroacetamide in D-glucose or sucrose medium. This is due to derepressed expression of the genes coding for acetyl CoA synthetase and acetamidase which convert these compounds to the toxic product fluoroacetate (Hynes and Kelly 1977). Selection procedures have used media containing D-glucose or sucrose together with the toxic analogues of acetamide, acetate and ethanol (fluoroacetamide, fluoroacetate and allyl alcohol respectively). This enabled the identification of mutations in which the synthesis of the enzymes involved in converting these compounds to toxic metabolites are no longer derepressed in D-glucose or sucrose media and possibly even more sensitive to regulation by carbon catabolite repression than wildtype strains (Kelly and Hynes 1977).

Finally, one other selection method resulted in the isolation of a mutant affected in carbon catabolite repression. \textit{frA} mutants cannot utilize D-fructose, D-sorbitol, D-mannitol or D-mannose as carbon sources, leading to inhibited growth on these compounds in the presence of D-glucose (reviewed by McCullough \textit{et al.} 1977). The reasons for \textit{frA} mutants leading to this toxicity remain undefined. \textit{creA} \textsuperscript{d-30} was isolated as a spontaneous mutation in a \textit{frA-1} strain which conferred resistance to 1\% D-mannitol in minimal medium with 1\% L-arabinose as the carbon source, although the basis for this
phenotype is not understood (Arst et al. 1990).

These selection methods have so far identified four loci involved in carbon catabolite repression in *A. nidulans* which have been designated creA, creB, creC and cre-34. creA has been mapped to linkage group I (Bailey and Arst 1975) while creB, creC and cre-34 are located on linkage group II with close linkage between creC and cre-34 (Hynes and Kelly 1977, Kelly and Hynes 1977).

1.5.2 The phenotype of creA mutants:

Mutations in the creA gene have been the most common mutations identified as leading to an altered response to carbon catabolite repression. All known creA mutant alleles have a partially derepressed phenotype regardless of the presence of D-glucose. That is, they result in a failure to respond normally to a source of carbon catabolite repression. The reported creA mutant alleles are all recessive to the wildtype allele in diploid strains. So far no creA alleles leading to the permanent repression of synthesis of enzymes involved in the utilization of alternative carbon sources have been identified even though efforts have been made to do so using a variety of potential selection procedures (Arst and Bailey 1977).

creA\textsuperscript{d-1} and creA204 have been shown to have similar phenotypic affects and were isolated as suppressors of the areA mutant phenotype by their inability to utilize L-proline and acetamide respectively in the presence of D-glucose (Arst and Cove 1973, Hynes and Kelly 1977). creA\textsuperscript{d-1} leads to partial derepression of alcohol dehydrogenase, proline oxidase and permease, sorbitol dehydrogenase,
mannitol dehydrogenase, NAD-linked glutamate dehydrogenase, galactokinase and/or galactose permease and acetamidase, while remaining unaffected for the utilization of either repressing or derepressing carbon sources (Arst and MacDonald 1975, Bailey and Arst 1975, Arst and Bailey 1977). The phenotypic effects of creA<sup>d-1</sup> in combination with a number of other mutations have been tested by Arst and Bailey (1977). The pycA-3 mutation leads to the loss of pyruvate carboxylase activity (Skinner and Armitt 1972) and therefore strains containing this mutation have a requirement for a source of tricarboxylic acid cycle intermediates. In a pycA-3:creA<sup>d-1</sup> double mutant, creA<sup>d-1</sup> allows pycA-3 to be at least partially supplemented by (+)-tartrate, acetamide, ethanol, L-ornithine and L-proline in the presence of D-glucose. This supplementation is prevented by D-glucose in pycA-3 single mutants, while in the double mutant carbon catabolite repression of the enzymes converting these compounds to TCA cycle intermediates is relieved (Bailey and Arst 1975).

Some mutants defective in sugar uptake have been shown to be resistant to the toxic effects of L-sorbose and 2-deoxy-D-glucose (Elorza and Arst 1971). However, creA<sup>d-1</sup> strains remain sensitive to these compounds and are unaffected in the uptake of D-glucose-<sup>14</sup>C (Bailey and Arst 1975).

Phenotypic variation exists among the creA mutant alleles in their abilities to suppress both areA and pdhA mutant phenotypes for growth on various compounds (Arst and Bailey 1977). creA<sup>d-2</sup> leads to derepressed levels of proline oxidase and proline permease while carbon catabolite repression of acetamidase is unaffected in the presence of a strong source of carbon catabolite repression. However, creA<sup>d-25</sup> results in the derepression of acetamidase but not of the
enzymes for L-proline metabolism. In creA<sup>d-1</sup> strains on the other hand, both acetamidase and enzymes for L-proline metabolism are derepressed. Therefore, creA<sup>d-1</sup> can suppress the areA mutant phenotype on glucose medium containing either acetamide or L-proline as the sole nitrogen source. Variation is also demonstrated by the ability of creA mutants to suppress the pdhA mutant phenotype. creA<sup>d-1</sup> and creA<sup>d-2</sup> allow considerable utilization of ethanol by pdhA strains in the presence of D-glucose while creA<sup>d-25;pdhA</sup> double mutants show poor supplementation by ethanol in the same medium. Therefore, creA<sup>d-1</sup> and creA<sup>d-2</sup> lead to greater relief of carbon catabolite repression of alcohol dehydrogenase than does creA<sup>d-25</sup>. These studies demonstrate that creA mutations do not necessarily affect the derepression of synthesis of the same enzymes subject to carbon catabolite repression, implying a non-hierarchical heterogeneity within this group of mutants (Arst and Bailey 1977). Such mutants cannot simply be null mutations, since all the mutants would be expected to have the same phenotype.

The extent to which various carbon sources are repressing or derepressing have been scored by the ability of areA and pdhA mutants to be supplemented by nitrogen sources and sources of acetyl CoA respectively in the presence of a range of carbon sources. This has shown that D-glucose, sucrose and D-xylose are strong sources of carbon catabolite repression, while L-arabinose, lactose, melibiose, meso-inositol and glycerol act as derepressing carbon sources in A. nidulans. Intermediate between carbon sources that are repressing and those that are derepressing, are D-mannose, D-galactose, D-sorbitol, D-fructose, D-mannitol and ethanol (Arst and Bailey 1977).

creA mutant alleles lead to a very compact colony morphology. As
with the variation found for other phenotypes of creA alleles, the affect on colony morphology is also variable between alleles. creA<sup>d</sup>-30 has the most extreme affect and leads to a severely compact morphology (Arst et al. 1990). creA<sup>d</sup>-2 also has an extreme affect on colony morphology while creA<sup>d</sup>-1 and creA204 strains have less affect. creA225 on the other hand, shows little affect on colony morphology (Arst and Bailey 1977, Hynes and Kelly 1977). Altered morphology has been reported in some yeast strains having defective carbon catabolite repression (Denis and Malvar 1990, Gosh et al. 1973, Montencourt et al. 1973, Sakai et al. 1990) and also in other organisms such as the crisp and frost mutants of N. crassa which are associated with altered cAMP levels within cells (Terenzi et al. 1974, Scott 1976). Arst and Bailey (1977) have suggested that altered regulation by carbon catabolite repression in A. nidulans may lead to altered morphology due to the defective synthesis of cell wall and membrane components such as carbohydrates, glucoproteins or glycolipids.

Extracellular protease activity is subject to carbon catabolite repression in wildtype A. nidulans and this activity can be assessed by the clearing of casein in powdered milk in solid medium (Cohen 1972, 1973). Therefore, cre mutants clear casein in the presence of D-glucose or sucrose whereas wildtype strains do not (Hynes and Kelly 1977).

1.5.3 The phenotype of creB and creC mutants:

creB15 and creC27 were isolated and characterized by Hynes and Kelly (1977) as affecting carbon catabolite repression by their
ability to suppress the phenotype of areA217 (Hynes 1975a) on 1% sucrose medium using acetamide as the sole nitrogen source and subsequently this suppression was shown not to be allele specific. Mutations in creB and creC are recessive and lead to the poor utilization of a large number of carbon sources such as L-proline, L-alanine, gluconate, D-glucuronate, L-arabinose, D-quinate, lactose, succinate, L-tyrosine, L-glutamate, L-glutamine, L-asparagine, L-ornithine, L-arginine, L-aspartate, D-fructose, L-rhamnose and D-mannose. While creB15 and creC27 strains both show reduced growth on these carbon sources, the creB15 phenotype being more extreme, creB and creC mutants differ in their ability to utilize other carbon sources. For example, creB15 leads to reduced utilization of D-mannitol and pyruvate as sole carbon sources, while creC27 strains are indistinguishable from wildtype A. nidulans for growth on these compounds. The utilization of other carbon sources such as sucrose, glucose, glycerol, acetate and ethanol is not affected in strains containing creB15 and creC27. In creB and creC mutant strains the enzymes found at altered levels in D-glucose/sucrose grown mycelia are closely related. Namely they are all enzymes of 2-carbon metabolism and include acetyl coA synthetase, acetamidase, alcohol dehydrogenase, fructose 1-6-diphosphatase, isocitrate lyase, malate synthase, NADP-1-isocitrate dehydrogenase and extracellular protease. In contrast to the creA alleles, the phenotypes for the creB and creC mutants are not restricted to derepression phenotypes. The synthesis of some enzymes such as alcohol dehydrogenase are partially derepressed in creB and creC mutants while the levels of other enzymes are found to be reduced without their sensitivity to carbon catabolite repression being altered as is the case with D-quinate dehydrogenase. Extracellular protease levels are found to be elevated.
regardless of the presence or absence of glucose in creB and creC strains as they are in creA mutant strains, while still other enzymes such as α-glucosidase show an increase in the derepressed levels of synthesis. creB and creC mutant alleles, like creA alleles, do not affect the uptake or metabolism of D-glucose itself but there is evidence that they do lead to a reduced uptake of L-glutamate and L-proline. creB15 and creC27 also show hypersensitivity to fluoroacetate and fluoroacetamide and sensitivity to allyl alcohol in glucose medium. This is also similar to the phenotype of creA mutant alleles on these media (Hynes and Kelly 1977, Kelly and Hynes 1977, Kelly 1980).

One other locus has been identified by the cre-34 mutation as being involved in carbon catabolite repression in A. nidulans. cre-34 was isolated by Kelly and Hynes (1977) as a spontaneous mutation in a creC27 strain, reversing the sensitivity of creC27 to fluoroacetamide in glucose medium. cre-34 was mapped to within 3-6 map units of creC27 and is recessive to wildtype. Further to this, it also relieves sensitivity to fluoroacetamide in the presence of a strong source of carbon catabolite repression and decreases growth on acetamide, L-proline and L-glutamate as nitrogen sources in creC27 and creA204 strains. cre-34 results in tighter carbon catabolite repression by sucrose of acetyl CoA synthetase and isocitrate lyase as compared to wildtype strains and lower levels of these enzymes in the presence but not absence of sucrose or glucose. cre-34 leads to the reversal of the allyl alcohol sensitivity in glucose medium of creC27 and affects utilization of L-proline, L-glutamate and milk powder clearing in the presence but not in the absence of glucose. Kelly and Hynes (1977) point out that these properties suggest that
some aspect of the utilization of these compounds is subject to
greater carbon catabolite repression in cre-34 strains.

1.5.4 Proposed roles for creA, creB, creC and cre-34 gene products:

On the basis of the observed effects of mutations in creA, creB and creC and cre-34 tentative roles for the products of these genes have been proposed.

Bailey and Arst (1975) suggested that creA probably codes for a negatively acting wide domain regulatory molecule involved in carbon catabolite repression, where a normal functional creA product is required for the repression of genes involved in the utilization of alternative carbon sources. That creA is a wide domain regulatory gene is suggested by virtue of the number and range of enzymes that are no longer subject to carbon catabolite repression in creA mutant strains, and that these alleles show phenotypic heterogeneity. The creA gene product is proposed to be negatively acting since all known mutant alleles are recessive and have the phenotype of derepression. If these alleles are loss of function alleles then mutants leading to permanent repression would be expected to be rare in a negatively acting regulatory system. Although dominant or semi-dominant mutations leading to failure to derepress the synthesis of a variety of enzymes in the absence of a source of carbon catabolite repression would be expected to occur rarely, extensive searches have not identified any showing this phenotype (Arst and Bailey 1977).

The creB and creC gene products have been shown to be more specific to the control of enzymes for the utilization of 2-carbon carbon sources. The lack of phenotypic heterogeneity among alleles
isolated (Kelly, pers. comm.) implies that creB and creC have a more specific role in carbon catabolite repression than does the creA gene product. On the tentative assumption that the creB and creC mutant alleles are loss of function alleles, then these gene products may also act negatively to control the synthesis of some enzymes. However, their role is unclear since the phenotype of creB and creC mutants is only partial derepression together with altered levels of activity found for other enzymes. Positive regulation by these gene products is compatible with the recessive nature of their poor utilization of some carbon sources (Kelly and Hynes 1977, Kelly 1980).

Kelly (1980) proposed that the phenotype and recessive nature of the cre-34 mutation suggests a positive role for the cre-34 gene product in carbon catabolite repression, probably in the presence but not absence of a source of carbon catabolite repression. This is supported by the fact that enzyme synthesis is not affected in the cre-34 mutant in the absence of glucose or sucrose.

1.5.5 Molecular studies of regulation by carbon catabolite repression in A. nidulans:

Only a limited amount of information is available to indicate how creA acts at the molecular level to regulate the large number of genes that are subject to carbon catabolite repression. Although the cloning and preliminary characterization of creA are outlined in this thesis, detailed molecular studies of the genes controlled by creA are also required in order to fully understand the molecular basis for regulation by creA. Molecular studies have begun to determine the
mechanism by which creA acts to control the regulation of other genes such as those involved in the utilization of acetamide, ethanol and L-proline as carbon sources.

The utilization of ethanol by A. nidulans is controlled by induction via the positively acting regulatory gene alcR and by glucose repression. The two structural genes required for ethanol metabolism, alcA encoding alcohol dehydrogenase and aldA encoding aldehyde dehydrogenase are subject to positive control by the alcR gene (Pateman et al. 1983, Sealy-Lewis and Lockington 1984). These three genes have been cloned and sequenced (Lockington et al. 1985, Gwyne et al. 1987, Pickett et al. 1987, Felenbok et al. 1988). The expression of alcR is subject to carbon catabolite repression and was found to be derepressed in the creA<sup>d</sup>-1 strain (Lockington et al. 1987). It has been demonstrated that carbon catabolite repression via creA acts independently on alcA and aldA. Evidence for this comes from the finding that these genes remain subject to carbon catabolite repression in transformants containing a construct of the alcR coding region under the control of the promoter of the glyceraldehyde-3-phosphate dehydrogenase gene which is not subject to carbon catabolite repression, and hence leads to derepressed expression of the alcR protein (Felenbok et al. 1989). Therefore, the creA protein independently controls the expression of both the structural genes and the positively acting regulatory gene alcR involved in the utilization of ethanol as a carbon source. Felenbok (1989) points out that since these three genes are subject to control by the creA protein then common sequences in their promoter regions may be expected to occur which identify a creA specific target site. No putative sites from sequence comparisons have been identified.
(Felenbok et al. 1988). However, other possible sequences which may represent creA target sites have been suggested in the promoter regions of \(\text{alcR}, \ \text{prnB}, \ \text{qutB}\) and \(\text{amdS}\) which are all subject to regulation by creA. Deletion of a 13bp sequence at position -136 in the upstream region of \(\text{alcR}\) has been carried out and this construct used to transform wildtype \(A. \ \text{nidulans}\). Initial reports of transformants possibly showing derepressed levels of \(\text{alcR}\) expression exist (Felenbok et al. 1989).

The regulation of the \(\text{amdS}\) gene encoding acetamidase in \(A. \ \text{nidulans}\) has been extensively studied by both classical and molecular techniques. A large number of 5' mutations that affect \(\text{amdS}\) expression have been sequenced and this approach has been used to identify target sites within the \(\text{amdS}\) control region for some of the proteins which regulate the expression of the gene (Hynes et al. 1988, reviewed by Davis and Hynes 1989). However, if creA acts directly on the \(\text{amdS}\) promoter region to mediate regulation by carbon catabolite repression, its target site remains to be identified.

O'Connell (1990) has investigated the upstream sequences of genes regulated by carbon catabolite repression from a number of filamentous fungi for putative creA-like target sites. It was found that a core sequence of 7bp has sequence similarity in the upstream control regions of a number of genes including \(\text{aldA}\) from \(A. \ \text{niger}\), \(\text{alcR}\), \(\text{alcA}\), \(\text{acuD}\), \(\text{acuE}\), \(\text{amdS}\), \(\text{facA}\), \(\text{prnB}\) and \(\text{qutE}\) from \(A. \ \text{nidulans}\) and \(\text{qa-2}\) from \(N. \ \text{crassa}\) which are all regulated by carbon catabolite repression. The significance of this sequence remains to be confirmed since it was also found (although with generally less overall homology) to be present in the upstream regions of other genes not subject to regulation by carbon catabolite repression. Furthermore,
deletion analysis of this sequence from the upstream region of \textit{aldA} from \textit{A. niger} was shown not to affect the expression of the gene (O'Connell 1990).

The mechanisms by which the products of the \textit{creA}, \textit{creB} and \textit{creC} genes act at the molecular level to regulate the large number of genes subject to carbon catabolite repression remain to be studied and understood. Biochemical and genetical studies of mutants with an altered response to carbon catabolite repression have suggested functions for these genes. This information must be coupled with the investigation of these genes and their products at the molecular level in order to begin to explain the mechanism of carbon catabolite repression in \textit{A. nidulans}.

1.6 AIMS OF THIS STUDY:

This thesis outlines the characterization of the \textit{creA} gene with a view to determining how it may be involved in carbon catabolite repression. Previous work reported on \textit{creA} suggested that it is one of the major genes involved in carbon catabolite repression in \textit{A. nidulans} and for this reason the study of this gene was chosen as a starting point to an understanding of carbon catabolite repression in \textit{A. nidulans}. This study set out to isolate and clone the \textit{creA} gene from \textit{A. nidulans} and to determine as far as possible, from the characterization of the gene, how \textit{creA} may act at the molecular level. It was of interest to determine whether the putative product of the \textit{creA} gene showed sequence similarities to negative regulators in support of \textit{creA}s negative role as suggested by genetical studies. This is in contrast to most regulators of transcription in \textit{A.
**nidulans** and other eukaryotes, which are generally positively acting. At this stage it cannot be ruled out that creA acts positively in that it may activate other molecules that derepress carbon catabolite repression, or general transcription factors that activate expression. However, analysis of the putative creA protein outlined in this thesis shows that it contains a domain similar to that found in other proteins known to be required for transcriptional repression function, further suggesting that the creA protein acts negatively in carbon catabolite repression.
CHAPTER TWO
MATERIALS AND METHODS.

2.1 MATERIALS.

General reagents: General chemicals, media requirements and supplements were of laboratory grade and were purchased from Aldrich Chemical Company Inc., BDH Ltd., Sigma Chemical Company, Oxoid Ltd. and United States Biochemical Corp.. Antibiotics, isopropylthiogalactoside (IPTG), 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and nuclease free bovine serum albumin were purchased from Boehringer-Mannheim GmbH. Bleomycin was a gift from M. J. Hynes.

Nucleotides and isotopes: Unlabelled deoxyribonucleotides and dideoxyribonucleotides were purchased from Boehringer-Mannheim GmbH. α-32P-dCTP and α-32P-dATP (3000 Ci/mnmole) and γ-32P-dATP (4000 Ci/mnmole) were purchased from BRESATEC Pty. Ltd.. Oligonucleotides for DNA sequencing and primer extension analysis were purchased from the Department of Microbiology, University of Adelaide.

Enzymes: Restriction endonucleases, calf intestinal phosphatase, polynucleotide kinase, sequencing grade klenow fragment of E. coli DNA polymerase I and proteinase K were purchased from Boehringer-Mannheim GmbH. Bacteriophage T4 DNA ligase and E. coli DNA polymerase
I were purchased from BRESATEC Pty. Ltd. M-MLV reverse transcriptase was purchased from Bethesda Research Laboratories Life Technologies Inc. Lysozyme, β-glucuronidase, ribonuclease A and pronase were purchased from Sigma Chemical Company. Novozyme was purchased from Novo Industries.

2.2 BUFFERS AND STOCK SOLUTIONS:

2.2.1 1 X Reaction buffers:

Restriction endonucleases: Reaction buffers for restriction enzyme digests were used as supplied by Boehringer-Mannheim GmbH.

Ligation buffer: 10mM MgCl₂, 10mM DTT, 1mM spermidine, 1mM ATP, 50mM Tris.HCl (pH 7.5).

Calf intestinal phosphatase buffer: 50mM Tris.HCl (pH 9.0), 1mM MgCl₂, 0.1mM ZnCl₃, 1mM spermidine.

Polynucleotide kinase buffer: 50mM Tris.HCl (pH 7.5), 10mM MgCl₂, 5mM DTT.

Nick translation buffer: 50mM Tris.HCl (pH 7.5), 7.5mM Mg(CH₃COO)₂, 4mM DTT, 100µg/ml BSA.

HIN buffer (C-tests): 10mM Tris.HCl (pH 7.4), 10mM MgCl₂, 50mM NaCl.

TM buffer (DNA sequencing): 10mM Tris.HCl (pH 8.0), 10mM MgCl₂.
Stock solutions:

1 X Denhardts: 0.02% (w/v) ficoll, 0.02% (w/v) polyvinlyridone, 0.02% (w/v) BSA (Pentax Fraction V).

Aspergillus salt solution: (per litre) 26g KCl, 26g MgSO₄.7H₂O, 76g KH₂PO₄, 50ml Aspergillus trace element solution, 2ml CHCl₃ as preservative.

1 X Loading buffer: 0.042% (w/v) bromophenol blue, 0.042% (w/v) xylene cyanol, 6.67% sucrose.

2 X M9 Salts: (per litre) 12g Na₂HPO₄, 6g KH₂PO₄, 1g NaCl, 2g NH₄Cl.

SM buffer: 100mM NaCl, 10mM MgSO₄.7H₂O, 50mM Tris.HCl (pH 7.5), 0.1% (w/v) gelatin.

1 X SSC: 0.15M NaCl, 0.15M Na₃C₆H₅O₇.2H₂O, pH 7.2.

1 X SSPE: 0.18M NaCl, 10mM NaH₂PO₄, 1mM EDTA, pH 7.4.

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

1 X TAE: 40mM Tris, 20mM NaCH₃COO, 2mM EDTA, pH 7.8.

1 X TBE: 89mM Tris, 89mM H₃BO₄, 2mM EDTA, pH 8.4.

1 X TE: 10mM Tris.HCl (pH 8.0), 1mM EDTA.
Aspergillus trace element solution: (per litre) 40mg Na$_2$B$_4$O$_7$, 400mg CuSO$_4$, 1g FePO$_4$, 600mg MnSO$_4$.H$_2$O, 800mg Na$_2$MoO$_4$.2H$_2$O, 8g ZnSO$_4$.7H$_2$O, 2ml CHCl$_3$ as preservative.

Aspergillus vitamin solution: (per litre) 40mg p-aminobenzoic acid, 50mg aneurin.HCl, 1mg D-biotin, 400mg inositol, 100mg nicotinic acid, 200mg calcium D-pantothenate, 100mg riboflavin, 50mg pyridoxine, 2ml CHCl$_3$ as preservative.

2.3 MEDIA:

2.3.1 Bacterial and yeast growth media:

L-broth: 1% (w/v) NaCl, 0.5% (w/v) yeast extract, 1% (w/v) tryptone, pH 7.5. Plates were solidified with 1.5% (w/v) agarose or Oxoid Class I agar.

M$_9$ glucose minimal medium: 50% 2 X M$_9$ Salts, 10mM MgSO$_4$, 1mM CaCl$_2$, 10mM thiamine.HCl, 0.2% (w/v) D-glucose, 50% (v/v) {3% (w/v) Oxoid Class I agar}.

SOC: L-broth plus: 1% (w/v) Oxoid tryptone, 1.6% (w/v) yeast extract, pH 7.5.

YEPD: 1% yeast extract (w/v), 2% (w/v) peptone, 2% D-glucose. Plates were solidified with 1% Oxoid Class III agar.

2 X YT: 0.5% (w/v) NaCl, 1% tryptone, 1.6% (w/v) yeast extract, pH 7.5.
2.3.2 *Aspergillus* growth media:

**Acetate medium:** 2% (v/v) *Aspergillus* salt solution, 50mM NaAc, pH 6.5. Plates were solidified with 1% (w/v) Oxoid Class I agar.

**Carbon free medium:** 2% (v/v) *Aspergillus* salt solution, pH 6.5. Plates were solidified with 1% or 2.2% (w/v) Oxoid Class I agar. Unless otherwise indicated, carbon sources were added to a final concentration of 1%.

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

**Nitrogen free medium (glucose medium):** 2% (v/v) *Aspergillus* salt solution, 1% (w/v) D-glucose, pH 6.5. Plates were solidified with 1% or 2.2% (w/v) Oxoid Class I agar. Unless otherwise indicated, nitrogen sources were added to 10mM.

**Protoplast medium:** 1M sucrose, 1% (w/v) D-glucose, 2% (v/v) *Aspergillus* salt solution, pH 7.0 and solidified with 0.25% or 1% (w/v) Oxoid Class I agar for overlayer and underlayer of plates respectively.

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

L-arginine 0.12mg/ml
D-biotin 0.01ug/ml
nicotinic acid 1.0ug/ml
p-aminobenzoic acid 50.0ug/ml
pyridoxal hydrochloride 0.5ug/ml
riboflavin 2.5ug/ml
sodium thiosulphate 0.1% (w/v)

Antibiotics: Antibiotics were added to media to the following concentrations:

blenoxane 1.0ug/ml
acriflavin 0.001% (w/v)
hygromycin 600ug/ml

2.4 E. coli STRAINS AND BACTERIOPHAGE:

E. coli strain DH1 (Hanahan 1983) was used for general plasmid maintenance. E. coli strain JM101 was used to screen recombinant plasmids using pUC19 as a vector and for propagating λM13 bacteriophage (Yannisch-Perron et al. 1985). Recombinant λZAP and λZAPII (Short et al. 1988) bacteriophage were propagated on E. coli strain NM538 (Frischauff et al. 1983) and recombinant λgt10 bacteriophage were propagated on E. coli strain C600 (Huynh et al. 1985). Aspergillus nidulans cDNA libraries in λZAP, λZAPII and λgt10 were obtained from W.E. Timberlake (Dean and Timberlake 1989). The titers for the libraries were determined to be $5.0 \times 10^8$, $1.8 \times 10^8$ and $2.5 \times 10^7$ respectively.
2.5 FUNGAL STRAINS:

A list of *Aspergillus* strains used in this study and their genotypes are given in Table 2.5.1. For meanings of gene symbols see Clutterbuck (1984).

Other fungal species used were *Penicillium chrysogenum* obtained from S. Andrews (S.A College of Advanced Education), *Neurospora crassa* strain T391 obtained from D. Catcheside (Flinders University, South Australia), *Saccharomyces cerevisiae* genotype a/α leu2; ura3-s2/ura3; ade1 obtained from P. Langridge (Waite Institute, South Australia) and strain 2289 99 of *Melampsora lini* obtained from J. Timmis (University of Adelaide, South Australia).

2.6 PLASMIDS:

A list of plasmids used in this study is shown in Table 2.6.1.

2.7 METHODS:

Unless otherwise stated, methods were carried out as described in the references cited.

2.7.1 Genetic manipulation and growth testing of *Aspergillus*:

Growth testing and meiotic analyses of *A. nidulans* were carried out as in Cove (1966). Diploids were haploidized on 1% complete medium (Hastie 1970) containing 1μl/ml of 0.075% benlate to select for haploids, unless otherwise specified. *A. nidulans* strain
Table 2.5.1: Genotypes of *Aspergillus* strains.

<table>
<thead>
<tr>
<th>STRAIN</th>
<th>GENOTYPE</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. nidulans</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wildtype</td>
<td><em>bia1</em>; <em>niiA4</em></td>
<td>Cove and Pateman (1963)</td>
</tr>
<tr>
<td>MSF</td>
<td><em>yA1</em> <em>ade20</em> <em>suAde201</em>; <em>acrA1</em>; <em>galE1</em>;</td>
<td>Kafer (1961)</td>
</tr>
<tr>
<td></td>
<td><em>pyroA4</em>; <em>facA303</em>; <em>sB3</em>; <em>nicB8</em>; <em>riboB2</em></td>
<td></td>
</tr>
<tr>
<td>J7</td>
<td><em>pabaA1</em>; <em>prn Δ353</em></td>
<td>Sharma and Arst (1985)</td>
</tr>
<tr>
<td>1070</td>
<td><em>bia1</em> <em>croAΔ-1</em></td>
<td>Bailey and Arst (1975)</td>
</tr>
<tr>
<td>664</td>
<td><em>bia1</em> <em>croA204</em>; <em>niiA4</em></td>
<td>Hynes and Kelly (1977)</td>
</tr>
<tr>
<td>SA20</td>
<td><em>yA1</em> <em>croA220</em>; <em>riboB2</em>; <em>areA217</em></td>
<td>Hynes (unpublished)</td>
</tr>
<tr>
<td>SA25</td>
<td><em>yA1</em> <em>croA225</em>; <em>riboB2</em>; <em>areA217</em></td>
<td>Hynes (unpublished)</td>
</tr>
<tr>
<td><em>croAΔ-30</em></td>
<td><em>bia1</em> <em>croAΔ-30</em></td>
<td>Arst et al. (1990)</td>
</tr>
<tr>
<td>C43</td>
<td><em>bia1</em> <em>pabaA1</em> <em>croA204</em>; <em>argB2</em></td>
<td>see text</td>
</tr>
<tr>
<td>C61</td>
<td><em>yA1</em>; <em>nicB8</em>; <em>pyroA4</em>; <em>niiA4</em> <em>riboB2</em></td>
<td>see text</td>
</tr>
<tr>
<td>C62</td>
<td><em>Acra1</em>; <em>galE1</em>; <em>facA303</em>; <em>sB3</em>; <em>bia1</em>; <em>riboB2</em></td>
<td>see text</td>
</tr>
<tr>
<td><em>A. niger</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wildtype</td>
<td>no mutant markers known</td>
<td>Kelly and Hynes (1985)</td>
</tr>
<tr>
<td><em>A. oryzae</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wildtype</td>
<td>no mutant marker known</td>
<td>S. Andrews (pers comm.)</td>
</tr>
</tbody>
</table>
Table 2.6.1: Plasmids used in this study.

<table>
<thead>
<tr>
<th>PLASMID</th>
<th>INSERT</th>
<th>VECTOR</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>pM006</td>
<td>argB</td>
<td>pUC19</td>
<td>Upshall (1986)</td>
</tr>
<tr>
<td>p3SR2</td>
<td>amdS</td>
<td>pBR322</td>
<td>Hynes et al. (1983)</td>
</tr>
<tr>
<td>pPL3</td>
<td>riboB</td>
<td>pUC19</td>
<td>Oakley et al. (1987a)</td>
</tr>
<tr>
<td>pAmPh366-1</td>
<td>$P_{m}^{r}$</td>
<td>pUCB</td>
<td>Austin et al. (1990)</td>
</tr>
<tr>
<td>pAN7-1</td>
<td>$Hyg^{r}$</td>
<td>pUC19</td>
<td>Punt et al. (1987)</td>
</tr>
<tr>
<td>pANC1</td>
<td>argB, creA</td>
<td>pUC19</td>
<td>see text</td>
</tr>
<tr>
<td>pANC2</td>
<td>argB</td>
<td>pUC19</td>
<td>&quot;</td>
</tr>
<tr>
<td>pANC3</td>
<td>creA (2.6kb)</td>
<td>pUC19</td>
<td>&quot;</td>
</tr>
<tr>
<td>pANC4</td>
<td>creA (2.3kb)</td>
<td>pUC19</td>
<td>&quot;</td>
</tr>
<tr>
<td>pANC5</td>
<td>creA (2.3kb)</td>
<td>pACYC184</td>
<td>&quot;</td>
</tr>
<tr>
<td>pANC6</td>
<td>creA (7.5kb)</td>
<td>pUC19</td>
<td>&quot;</td>
</tr>
<tr>
<td>pANC7</td>
<td>creA (7.5kb)</td>
<td>pUCBX</td>
<td>&quot;</td>
</tr>
<tr>
<td>pANC8</td>
<td>riboB + creA</td>
<td></td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>flanking regions</td>
<td></td>
<td>&quot;</td>
</tr>
<tr>
<td>pUCBX</td>
<td>pUC19 with the BamHI and XbaI</td>
<td></td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>sites deleted from the polylinker.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MSF was used as the multiply marked strain for haploidization analyses (Kafer 1961). Mutagenesis of specific strains was achieved by exposing conidial suspensions to UV (254.8nm) light, at a distance of 7cm for ten minutes.

2.7.2 Transformation of Aspergillus:

The isolation, preparation and transformation of *A. nidulans* protoplasts with plasmid DNA was by the method of Tilburn *et al.* (1983). ArgB+ transformants were selected as in Berse *et al.* (1983). Selection of AmdS+ transformants was as in Tilburn *et al.* (1983), RiboB+ transformants as described by Oakley *et al.* (1987a), hygromycin B resistant transformants as in Punt *et al.* (1987) and bleomycin resistant transformants as in Austin *et al.* (1990). The selection of CreA+ transformants is outlined in the text (Chapter 3, section 3.1.3).

2.7.3 Isolation of nucleic acids:

*Aspergillus, Neurospora* and *Penicillium* DNA was isolated by the method of Hynes *et al.* (1983) from freeze dried mycelia which had been harvested from overnight liquid cultures grown at 37°C. Specific size fractions of genomic DNA for the construction of libraries were recovered by spinning partial digests through sucrose step gradients (Griffith, 1979) or by electrophoresing digests in agarose gels of the appropriate concentration for separation, followed by squeeze-freezing the desired region of the gel to recover the DNA by the method of Thuring *et al.* (1975).
DNA from *Saccharomyces* was isolated from cultures grown at 30°C by the same method as that for *Aspergillus* DNA preparations.

For the isolation of *Aspergillus* RNA, cultures were grown overnight in liquid glucose media at 37°C, harvested and washed with warm carbon free medium and then transferred to appropriate prewarmed medium for four hours. Cultures were then harvested and washed with cold sterile distilled water and freeze dried overnight. Total RNA was isolated from the freeze dried mycelia as in Rienert *et al.* (1981).

For the large scale preparation of plasmid DNA, *E. coli* harbouring plasmids were cultured in L-broth plus 50μg/ml ampicillin and amplified using chloramphenicol to a final concentration of 170μg/ml. Cells were harvested by centrifugation at 3500g for 10 minutes, washed with 100mls 1 X STE and repelletted. Plasmid preparations were then made using either of the two following methods:

(A). Cells were resuspended in 10mls of 25% sucrose; 50mM Tris.HCl pH 8.3; 50mM NaCl; 10mM EDTA plus an extra 2.5mls of 0.5mM EDTA. Lysozyme was added to 2mg/ml and the cells incubated at room temperature for 30 minutes and a further 20 minutes on ice. 15mls of 2% Triton-X; 50mM Tris.HCl pH 8.3; 10mM EDTA was added and the precipitate pelleted for 90 minutes at 20,000 r.p.m. To the supernatant was added 40% PEG to a final concentration of 1M, 1M NaCl and 1mg of heat treated ribonuclease A and incubated overnight at 4°C. DNA was pelleted at 8,500g for 10 minutes, resuspended in 4mls 1 X TE and extracted once with chloroform:isoamyl alchol (24:1). Plasmid DNA was purified on CsCl equilibrium gradients by the method of Maniatis *et al.* (1982).
(B). Cells were resuspended in 4mls 15% sucrose; 25mM Tris.HCl pH 8.0; 10mM EDTA and incubated on ice for 40 minutes after the addition of 8mgs of lysozyme. 8mls of 0.2N NaOH; 1% SDS was added and the solution held on ice for 10 minutes followed by the addition of 5mls of 3M NaAc pH 4.6 and incubated on ice for a further 40 minutes. The precipitate was pelleted at 5,000g for 15 minutes and the supernatant incubated with 100ug heat treated ribonuclease A at 37°C for 40 minutes. This was followed by a phenol:chloroform (1:1) and one chloroform extraction. DNA was precipitated by the addition of 1/10 volume 3M NaAc pH 5.5 and 2.5 volumes of ethanol. The pellet was resuspended in 1.6mls of water and the plasmid DNA was precipitated by the addition of 0.4ml 13% PEG, incubated on ice for 1 hour and pelleted at 8,500g for 10 minutes. Plasmid DNA was resuspended in 1 X TE.

Small scale preparation of plasmid DNA was performed by the alkaline lysis method of Ish-Horowicz and Burke (1981).

Bacterial genomic DNA was isolated from overnight cultures. Cells were pelleted at 3500g for 10 minutes and incubated in 10mls 10mM Tris.HCl (pH 8.0), 10mM NaCl, 10mM EDTA, 0.5% SDS, 50ug/ml proteinase K per gram of cells for 30 minutes at 37°C. The lysed cells were then extracted three times with buffer saturated phenol. Two volumes of ethanol were added to the final aqueous phase and the precipitated DNA spooled out. The DNA was washed with 70% ethanol, resuspended in 1 X TE, RNAase treated, extracted with phenol and reprecipitated with ethanol.

DNA was prepared from lambda bacteriophage by the plate lysate method of Maniatis et al. (1982).
For the preparation of single stranded DNA for sequencing, single plaques of M13 bacteriophage were propagated in 2ml cultures of *E. coli* in 2 X YT medium as described by Ausubel *et al.* (1987). Cultures were centrifuged at 3,500g for 10 minutes and 1.5mls of the supernatant was collected in a microcentrifuge tube. This was spun at 12,000 r.p.m for 10 minutes and 1ml of the supernatant was decanted into a fresh microcentrifuge tube. 275ul of 20% PEG 8,000; 2.5M NaCl was added and the tubes incubated at room temperature for 15 minutes. Bacteriophage particles were pelleted by microcentrifugation at 12,000 r.p.m for 10 minutes. The supernatant was decanted and discarded. The tubes were spun for a further 2 minutes and the residual supernatant was removed with a drawn-out pasteur pipette. The bacteriophage were resuspended in 200ul of 1 X TE, extracted with 100ul of tris-buffered phenol and 150ul of the aqueous layer being collected. This was extracted with 75ul of chloroform and 100ul of the aqueous layer was collected. M13 DNA was precipitated by the addition of 11ul of 4M LiCl and 275ul of ethanol, incubated at -70°C for 15 minutes and collected by microcentrifugation at 12,000 r.p.m. Pellets were resuspended in 12ul of 1 X TE. For double stranded sequencing, 15ug of purified plasmid in 1 X TE was incubated for 15 minutes at 37°C with 5ul of 1N NaOH, 1mM EDTA. The sample was spin dialysed through sepharose CL-6B equilibrated in 1 X TE. 7ul of spin dialysate was used per sequencing reaction.

2.7.4 General recombinant DNA methods:

DNA manipulation and modification: General methods were used as indicated by the suppliers of enzymes or as outlined in laboratory manuals (Maniatis *et al.* 1982; Ausubel *et al.* 1987).
Transformation of *E. coli*: Transformation of *E. coli* was performed either by electroporation using a Bio-Rad Gene Pulser Version 1.0 and the methods of the supplier or by heat shocking as outlined in Maniatis *et al.* (1982). Cells were made competent for heat shock transformation by the following method: 50ml cultures in late log phase were pelleted for 2 minutes at 10,000 r.p.m., resuspended in 25mls 0.1M MgCl₂ and re-pelleted. The cells were resuspended in 25mls of 50mM CaCl₂ and incubated on ice for 20 minutes. This was followed by a final spin down and the cells resuspended for storage in a total of 5mls of 50mM CaCl₂, 15% glycerol. Cells were made competent for transformation by electroporation by the following procedure: 500ml cultures were grown to an O.D.₆₀₀ of 0.4, pelleted for 10 minutes at 5,000g and resuspended on ice in 250mls of cold 10% glycerol. Cells were repelleted and resuspended in 10mls of cold 10% glycerol twice and 5mls of cold 10% glycerol once. The final pellet was resuspended in 2.5mls cold 10% glycerol and the cells were stored at -80°C in 100ul aliquots. These cells were thawed on ice and incubated with DNA for 1 minute on ice prior to electroporation. 1ml of SOC medium was added to electroporated cells and incubated for 1 hour prior to plating out.

2.7.5 Gel electrophoresis:

Agarose gel electrophoresis of DNA was carried out in gels made from 1 X TAE buffer and an appropriate concentration of agarose. DNA was recovered from agarose gels by the squeeze-freeze method of Thuring *et al.* (1975). Bacteriophage lambda DNA digested with *HindIII* was used as the molecular weight marker.
RNA was electrophoresed in 1.5% (w/v) agarose, 8.0% formaldehyde, 10mM sodium orthophosphate (pH 7.0) gels with 10mM sodium orthophosphate (pH 7.0) as the running buffer. 10ug or 20ug samples of total RNA were prepared for electrophoresis by the addition of two volumes of {50% (v/v) formamide; 12% (v/v) formaldehyde; 10mM sodium orthophosphate (pH 7.0)} and heated for 10 minutes at 65°C.

DNA sequencing reactions were electrophoresed in gels containing 4-8% (w/v) polyacrylamide, 8M urea, 1 X TBE. Some sequences were also resolved by the addition of 25% (v/v) deionized formamide to the gel mix. The dimensions of the sequencing gels was 450 X 170 X 0.2mm and these were run at 1500 - 2000 volts.

2.7.6 \( ^{32}\)P Labelling of DNA:

DNA probes were prepared for hybridizations by nick translation. Specifically, reactions contained 250ng of DNA to be labelled, 0.1mM each of 3 dNTP's and 1ug/ml DNAase in 1 X nick translation buffer. This was incubated at room temperature for 30 minutes, heated to 65°C for 5 minutes and placed on ice. 20uCi \( \alpha^{32}\)PDNTP and 10 units of E. coli DNA polymerase I were added and the reaction carried out for 30 minutes at 16°C. Unincorporated nucleotides were removed using Bio-Rad Bio-Gel P60 columns.

Strand specific probes were created by annealing recombinant M13 clones to a reverse sequencing primer (25mer) obtained from BRESATEC Pty. Ltd. dATP, dGTP and dTTP were added to a final concentration of 0.5mM, together with 40uCi of \( \alpha^{32}\)-P-dCTP and one unit of Klenow. The reaction was extended for 30 minutes at 37°C and chased with 0.5mM
dNTP's for a further 15 minutes. Unincorporated nucleotides were removed using Bio-Rad Bio-Gel P60 columns. These probes were not denatured prior to hybridization to RNA.

Molecular weight markers were end labelled by combining up to 5μg of DNA with 0.08 mM each of three dNTP's, 10μCi α-32PdNTP and 1 unit of Klenow in 1 X low salt restriction enzyme buffer. The reaction was incubated at room temperature for 30 minutes and unincorporated nucleotides removed using Bio-Rad Bio-Gel P60 columns.

2.7.7 Nucleic acid blotting and hybridization conditions:

Southern blotting, Northern blotting and slot blotting of RNA and DNA using a Schleicher and Schuell minifold II apparatus was to Zeta-Probe membrane (Bio-Rad) by the alkali transfer methods recommended by the supplier. Plaque lifting to nitrocellulose was performed by the method of Benton and Davis (1977). Biodyne membrane (Pall) was used for colony blots and carried out by the SDS/NaOH method of Maniatis et al. (1982).

Hybridization of nick translated probes to Zeta-Probe membrane and washing of filters after hybridization was performed by the methods recommended by the supplier. The same conditions were used for colony blots on Biodyne except that 15μg/ml E. coli DNA was included in the prehybridization solution. For nitrocellulose, filters were prehybridized for 4 hours at 42°C using 50% formamide, 4 X SSPE, 10 X Denhardtts, 0.1% (w/v) SDS and 30μg/ml sonicated denatured salmon sperm DNA. Hybridization of denatured probes was under the same conditions except for the addition of 10% (w/v) dextran sulphate. Filters were washed to a final stringency of 0.1 X
SSC; 1.0% SDS at 65°C unless otherwise stated. Autoradiography was carried out using intensifying screens except for the autoradiography of quantitative slot blot analyses.

2.7.8 Quantification of DNA copy number and mRNA levels:

crel mRNA levels and DNA copy number were determined by hybridization of slot blots of serial dilutions of DNA and total RNA using the insert of pANC4. The quantity of DNA or RNA in each series was standardized using the argB gene of pMO06 or the riboB gene of pPL3 in the case of transformants. The signal on autoradiograms was quantified using a LKB Ultrascan XL enhanced laser densitometer. The absorbance units detected were plotted against the amount of RNA blotted to obtain data points within a linear range for the X-ray film. The arbitrary value of absorbance units/ug for the creA specific probe was corrected using the other probes and values were then normalized against samples indicated.

2.7.9 DNA sequencing:

The creA clone was subcloned into the bacteriophage vectors M13mp18 and M13mp19 for single stranded DNA sequencing using standard cloning techniques. Complementarity tests (c-tests) were carried out to determine the orientation of M13 clones by combining 5ul of two single stranded preparations and 8ul of 10 X HIN buffer, boiling for three minutes and leaving at room temperature for 30 minutes prior to electrophoresis. DNA sequencing of both plasmid and single stranded DNA was performed by the dideoxynucleotide chain termination method (Sanger et al. 1976) using Sequenase Version 2 kits (United
States Biochemical Corp.) or by the following procedure using Klenow: 2.5ng of Bio-Labs -40 sequencing primer or a specifically synthesized oligonucleotide was annealed to 8ul of template DNA in 1 X TM buffer by heating the mix to 100°C and then cooled slowly to room temperature. The annealed DNA was added to a tube containing 20uCi of α-<sup>32</sup>P-dCTP and 1 unit of Klenow. 2.5ul of this was aliquoted into four tubes containing one No (Table 2.7.1) solution (2.0ul) and 2.0ul of the corresponding ddNTP. The dideoxynucleotides ddATP, ddCTP, ddGTP and ddTTP were aliquoted from stock solutions of 0.3mM, 0.05mM, 0.15mM and 0.5mM respectively. The reactions were incubated at 37°C for 15 minutes and then 2.5ul of chase (0.5mM dCTP, 0.5mM dGTP, 0.5mM dTTP, 0.5mM dATP and 0.1 units/ul Klenow) was added to each tube and the reactions incubated for a further 15 minutes. 4ul of loading buffer (10mg/ml bromophenol blue, 10mg/ml xylene cyanol, 1mM EDTA in deionized formamide) was added to stop the reactions. Samples were denatured at 100°C for 10 minutes and 2ul was loaded onto the sequencing gel. After electrophoresis, the gel was fixed in {20% (v/v) methanol, 10% (v/v) acetic acid} for 20 minutes, dried and exposed to X-Ray film overnight. Compressions and pauses in the banding pattern were resolved by substituting dITP for dGTP in the reactions or by incubating reactions at 50°C or by the use of formamide in the gel mixture (see section 2.7.5).

Computer analysis of sequence data was carried out using Staden, NIH programs and the Genetics Computer Group Sequence Analysis Software Package Version 6.1 (Devereux 1984). Database comparisons were to the Genbank, EMBL, NBRF-Nucleotide and NBRF-Protein databases.
Table 2.7.1: dNTP solutions for DNA sequencing.

<table>
<thead>
<tr>
<th></th>
<th>A&lt;sup&gt;o&lt;/sup&gt;</th>
<th>T&lt;sup&gt;o&lt;/sup&gt;</th>
<th>G&lt;sup&gt;o&lt;/sup&gt;</th>
<th>C&lt;sup&gt;o&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5mM dATP</td>
<td>4.3ul</td>
<td>43ul</td>
<td>43ul</td>
<td>29ul</td>
</tr>
<tr>
<td>0.5mM dTTP</td>
<td>43ul</td>
<td>4.3ul</td>
<td>43ul</td>
<td>29ul</td>
</tr>
<tr>
<td>0.5mM dGTP</td>
<td>43ul</td>
<td>43ul</td>
<td>4.3ul</td>
<td>29ul</td>
</tr>
<tr>
<td>0.5mM dCTP</td>
<td>1ul</td>
<td>1ul</td>
<td>1ul</td>
<td>1ul</td>
</tr>
<tr>
<td>1 X TE</td>
<td>8.7ul</td>
<td>8.7ul</td>
<td>8.7ul</td>
<td>12ul</td>
</tr>
</tbody>
</table>
2.7.10 Primer extension analysis:

Primer extension analysis was carried out by procedure 1, for RNA sequencing, outlined in Geliebter et al. (1986) except that the ddNTP's were omitted from the transcription buffer. A specific oligonucleotide was end labelled with γ-32-P-dATP using polynucleotide kinase and annealed to RNA at a temperature determined by: annealing temp. in °C = 4(G+C)+2(A+T)-5.
CHAPTER THREE
THE CLONING OF \textit{creA} FROM \textit{Aspergillus nidulans}.

This chapter describes the cloning and the preliminary characterization of the \textit{creA} gene from \textit{Aspergillus nidulans}. A specific cloning strategy was chosen, previously shown to be useful for the isolation of regulatory genes, which have low levels of expression and are of unknown sequence. This strategy involved the transformation of a strain containing a \textit{creA} mutant allele with a genomic library, selection for complementation of the mutation and the physical rescuing of transforming sequences from the complemented transformant. Similar techniques have been employed by a number of workers to isolate regions containing both structural and regulatory genes of interest from \textit{A. nidulans} including the regulatory genes \textit{amdR} (Andrianopoulous and Hynes 1988), \textit{facB} (Katz and Hynes 1989a) and \textit{brlA} (Boylan \textit{et al.} 1987) and the structural genes \textit{yA} (Yelton \textit{et al.} 1985), \textit{acuD} (Ballance and Turner 1986), \textit{pyrG} (Oakley \textit{et al.} 1987a) and \textit{riboB} (Oakley \textit{et al.} 1987b). Evidence is presented in this section which indicates that the region which has been cloned from \textit{A. nidulans} contains the \textit{creA} locus.
3.1 THE CLONING OF creA:

3.1.1 The construction of a wildtype genomic library from A. nidulans:

DNA from wildtype A. nidulans was partially digested with MboI in reactions to optimize the number of fragments in the 6-15kb size range. Fragments of this size range were purified and ligated into the BamHI site of pM006. Ligation reactions were transformed into E. coli strain DHI. Plasmid harboring colonies were selected on medium containing ampicillin and pooled into four batches. Large scale preparations of plasmid library were made from cultures incubated for only three doublings without amplification. In total, these libraries represented 12,000 clones with an average insert size of 9.0kb and therefore there was a greater than 95% chance that a sequence of interest was represented among these clones. The number of clones required =

\[
\frac{\ln (1-\text{probability sequence being present})}{\ln \left(1-\frac{\text{average insert size}}{\text{size of genome}}\right)}
\]

3.1.2 The identification of a non-revertable creA- allele:

Allyl alcohol is toxic to A. nidulans when converted, via alcohol dehydrogenase, to its toxic product acrolein. Wildtype strains are resistant to allyl alcohol in the presence of D-glucose or sucrose due to carbon catabolite repression of alcohol dehydrogenase enzyme activity. However, creA mutations lead to complete sensitivity
to 2.5mM allyl alcohol in sucrose or D-glucose media because alcohol dehydrogenase activity is no longer subject to carbon catabolite repression (Hynes and Kelly 1977). Therefore, complementing transformants or revertants of a creA mutant strain may be identified by selection for resistance to allyl alcohol on D-glucose or sucrose medium. However, mutations in the structural gene for alcohol dehydrogenase, alcA, can also lead to a resistant phenotype. These mutants can be distinguished from wildtype strains by their resistance to allyl alcohol in glycerol medium (Ciriacy, 1975) and their poor utilization of ethanol as a carbon source. The extreme toxicity of acrolein in A. nidulans provides a powerful means of selecting for strains which do not have alcohol dehydrogenase enzyme activity when grown in the presence of D-glucose or sucrose. This provided a means by which it was possible to select for complementation of creA mutant alleles with the wildtype library on D-glucose or sucrose medium containing 2.5mM allyl alcohol. For this selection system to be successful in identifying true complementation of creA mutant alleles, it was important to determine the frequency with which the creA mutant alleles could phenotypically revert to wildtype on sucrose plus allyl alcohol medium.

Conidial suspensions of strains containing the mutant alleles creA<sup>d</sup>-1 or creA204 were either spread directly, or after UV mutagenesis, onto sucrose medium containing 2.5mM allyl alcohol. Colonies that grew (Figure 3.1.1) were screened for growth on ethanol minimal medium and those that could not utilize ethanol as a sole carbon source were classified as mutations affecting alcohol dehydrogenase activity. Colonies which could utilize ethanol as a carbon source, showed sensitivity to allyl alcohol in glycerol
Figure 3.1.1: Spontaneous and induced reversion of creA<sup>d</sup>-30 and creA204. *A. nidulans* strains 1070 and 664 are plated on sucrose media containing 2.5mM allyl alcohol. Conidial suspensions of each strain were plated out either directly (spontaneous) or after UV mutagenesis (induced) and grown for two days at 37°C in order to determine if these strains could phenotypically "revert" to CreA<sup>+</sup>. Strain 1070 showed high background growth and had a high "reversion" frequency on this medium.
medium, and were resistant to allyl alcohol on sucrose medium were classified as CreA+ "revertants" (Table 3.1.1). Strain 664, containing the creA204 allele did not show any spontaneous or induced phenotypic reversion to CreA+. This strain was crossed into an argB2 background (strain C26.1.10) to give strain C43 (biA1 pabaA1 creA204; argB2) and was used as the recipient strain for transformations with the plasmid libraries. The strain containing the creA<sup>d</sup>-1 allele (1070) produced many mutants affecting alcohol dehydrogenase activity and also potential CreA+ "revertants" when plated out in the same manner, on sucrose plus allyl alcohol medium. The relatively high "reversion" rates seen for creA<sup>d</sup>-1 was probably strain dependent, although this allele was not tested in a different genetic background. Since there was an increased frequency of both AlcA+ and CreA+ phenotypic classes, this did not necessarily indicate a highly revertable allele. Such high mutation rates could possibly have been due to a defective DNA repair system in strain 1070. Ten of these CreA+ "revertants" were chosen for further characterization. These phenotypic revertants segregated creA<sup>d</sup>-1 progeny when outcrossed to strain J7 which indicated that these strains contained phenotypic suppressors of the creA<sup>d</sup>-1 allele. Diploids were constructed between these ten strains and MSF, in an attempt to map some of these suppressors to a particular linkage group by haploidization analysis. Initially, some of these potential suppressors appeared to map to linkage group V or VII, however these strains seemed to be highly unstable and were found to spontaneously acquire mutations which could be characterized as Alc- or others that were shown to have poor utilization of ethanol as the sole carbon source. The latter group were not classified as Alc- since, unlike Alc- strains, they remained sensitive to allyl alcohol in the presence of glycerol. This
Table 3.1.1: Spotaneous and induced "reversion" of creA<sup>d</sup>-1 and creA<sub>204</sub>.

<table>
<thead>
<tr>
<th></th>
<th>creA&lt;sub&gt;204&lt;/sub&gt;</th>
<th></th>
<th>creA&lt;sup&gt;d&lt;/sup&gt;-1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spont.</td>
<td>Induced*</td>
<td>Spont.</td>
<td>Induced*</td>
</tr>
<tr>
<td>&quot;reversion&quot; freq.</td>
<td>1.3x10&lt;sup&gt;-6&lt;/sup&gt;</td>
<td>1.1x10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>1.4x10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>0.038</td>
</tr>
<tr>
<td>no. tested</td>
<td>40</td>
<td>46</td>
<td>37</td>
<td>48</td>
</tr>
<tr>
<td>no. Alc&lt;sup&gt;-&lt;/sup&gt;</td>
<td>40</td>
<td>46</td>
<td>33</td>
<td>40</td>
</tr>
<tr>
<td>no. CreA&lt;sup&gt;+&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

Conidial suspensions were plated onto sucrose minimal medium containing 2.5mM allyl alcohol and 10mM ammonium tartrate. Colonies were screened on 1% ethanol plus 10mM ammonium chloride. * Conidia were treated with UV, with a survival rate of 10%.
suggested that these strains had levels of alcohol dehydrogenase activity which were capable of converting a sufficient amount of allyl alcohol to acrolein to cause toxicity while remaining subject to carbon catabolite repression. The difficulties inherent in maintaining the stability of these strains did not allow for the full characterization of their genotypes as part of this work. This analysis made it clear that strain 664 containing the creA204 allele was more appropriate for this work.

3.1.3 Complementation of the creA204 mutation:

Strain C43 carrying the creA204 allele was transformed with preparations of the wildtype genomic libraries in pM006. Transformed protoplasts were either (a) plated onto minimal protoplast medium to select for arginine prototrophy and then screened for complementation of the creA204 mutation by replica plating to sucrose minimal medium plus 2.5mM allyl alcohol or, (b) plated onto minimal medium selecting for arginine prototrophy for ten hours and then overlayed with sucrose minimal medium plus 2.5mM allyl alcohol directly selecting for complementation of both argB2 and creA204.

In a total of eight experiments these selection techniques produced approximately one hundred and fifty colonies that grew on sucrose and allyl alcohol medium. Three of these colonies (which arose on the same transformation plate and subsequently were deemed to have originated from a single transformant) were able to utilize ethanol as a sole carbon source, and were therefore not alc- mutants. The remaining colonies showed poor utilization of ethanol as a carbon source and were resistant to allyl alcohol in glycerol medium, and
were classified as qic- mutants.

These complemented transformants (T41a, T41b and T41c) were tested for growth on a range of media to rule out the possibility that the observed phenotype was alcohol dehydrogenase specific (see Figure 3.1.2b for complementation for growth on sucrose plus allyl alcohol medium) by determining whether they also showed complementation for some of the other phenotypes due to creA mutant alleles. T41a, b and c showed wildtype morphology on complete medium (Figure 3.1.2a), unlike the compact morphology of creA mutant strains. While wildtype A. nidulans strains are sensitive, creA mutations lead to hypersensitivity to fluoroacetate and fluoroacetamide in glucose medium due to the derepressed expression of genes coding for the enzymes converting these compounds to fluoroacetate (Hynes and Kelly 1977). T41a, b and c grew slightly better than wildtype on glucose medium containing these compounds (Figure 3.1.2c and d), indicating repressed levels of acetyl-Co-A synthetase and acetamidase in the presence of a source of carbon catabolite repression. Extracellular protease activity is subject to carbon catabolite repression in wildtype A. nidulans and this activity can be assessed by the clearing of casein in powdered milk in solid media (Cohen 1972 and Cohen 1973). The creA204 mutation showed distinct clearing of milk in the presence of glucose whereas T41a, b and c showed lower levels of clearing (Figure 3.1.2e). Therefore preliminary growth testing showed that T41 was complemented for the creA204 allele of strain C43.
Figure 3.1.2: Growth testing of T41a, T41b and T41c.

Plates contain:

(A). Complete medium.
(B). Sucrose medium containing 2.5mM allyl alcohol.
(C). D-glucose medium containing 10mg/ml fluoroacetate.
(D). D-glucose medium containing 10mg/ml fluoroacetamide.
(E). D-glucose medium containing 1% milk powder.

(A), (B), (C) and (D):

Top left : wildtype A. nidulans.
Top right : creA204.
Bottom left to right : T41a, T41b and T41c.

(E):

Top left : creA204.
Top right : wildtype A. nidulans.
Bottom left to right : T41c, T41b and T41a.
3.1.4 Rescue of transforming sequences:

Genomic DNA prepared from T41 was partially digested with \textit{BamHI}, \textit{PstI} and \textit{XhoI} separately with the aim of generating fragments containing both pM006 vector sequences and the complementing sequence of interest. These partial digests were recircularized by ligation and transformed into \textit{E. coli} strain HB101. Colonies harbouring pM006 vector sequences were selected on medium containing ampicillin. Twenty ampicillin resistant colonies resulted from this experiment, of which nine, nine and two of these originated from partial digests using \textit{BamHI}, \textit{PstI} and \textit{XhoI} respectively. Preliminary restriction enzyme digest analysis of plasmids isolated from these colonies showed that six different plasmids were represented. Those which appeared to be the same were pooled for large scale plasmid preparations and used to re-transform \textit{A. nidulans} strain C43. Each plasmid preparation was tested for its ability, on transformation of strain C43, to either (A) complement both \textit{argB2} and \textit{creA204} by selecting for transformants on minimal medium for 10 hours and then overlayering with sucrose medium containing 2.5mM allyl alcohol or (B) complement \textit{creA} alone by plating transformed protoplasts onto media containing arginine and overlayering with sucrose medium plus allyl alcohol after 16 hours. One of the plasmid pools complemented both the \textit{argB2} and \textit{creA204} mutations (Figure 3.1.3). Eight of the twenty rescued plasmids were represented in this plasmid pool which complemented C43 for both mutations, of which all originated from religated \textit{BamHI} partial digests of T41 DNA. Further restriction digest analysis demonstrated that these 8 plasmids were identical. A restriction map of this plasmid, pANCl, is shown in Figure 3.1.4.
Figure 3.1.3: Complementation of C43 with pANC1.

Plates are:

Right: Control of untreated protoplasts plated onto protoplast medium with all supplements added except arginine and overlayed with protoplast medium containing 2.5 mM allyl alcohol.

Left top: pANC1 treated protoplasts plated onto protoplast medium with all supplements added except arginine. Many of these transformants appeared to turn yellow when grown for more than five days at 37°C. However, only green colonies were observed when treated protoplasts were plated at low density and only green transformants could be isolated when yellow regions were subcultured. The reason for the initial yellow appearance of some of these transformants remains unclear.

Left bottom: pANC1 treated protoplasts plated onto protoplast medium with all supplements added except arginine and overlayed with protoplast medium containing 2.5 mM allyl alcohol.

Therefore, pANC1 complemented C43 for both the creA204 and argB2 mutations.
Figure 3.1.4: Partial restriction map of pANC1. The regions of pANC1 that were subcloned into pUC19 to give pANC2, pANC3 and pANC4 are indicated.

Restriction site abbreviations:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>BamHI</td>
</tr>
<tr>
<td>Bg</td>
<td>BglII</td>
</tr>
<tr>
<td>E</td>
<td>EcoRI</td>
</tr>
<tr>
<td>S</td>
<td>SalI</td>
</tr>
<tr>
<td>X</td>
<td>XbaI</td>
</tr>
</tbody>
</table>
Southern hybridization analysis of XbaI cut wildtype DNA using pANC1 as a probe showed two bands of hybridization, one of 3.3kb corresponding to the native argB sequence and another of approximately 10.5kb corresponding to the cloned sequence. XbaI digested T41 DNA had two extra bands of approximately 3.7kb and 9.7kb indicating that transforming sequences had not integrated into either the argB native locus or the native site for the cloned sequence (Figure 3.1.5).

3.1.5 The localization of creA within pANC1:

Restriction fragments of pANC1 were subcloned into pUC19 and transformed into strain C43 in order to localize a smaller fragment capable of complementing the creA204 mutation alone. Transformation experiments using these plasmids were plated out, selecting separately for argB+ and creA+ transformants in each case. Southern analysis of pANC1 digests probed with pUC19 confirmed that the 2.9kb XbaI - EcoRI fragment of pANC1 represented pUC19 sequences. pANC2 contained the 3.0kb XbaI fragment of pANC1 (see Figure 3.1.4) and complemented the argB2 mutation when transformed into strain C43. pANC3 contained the 2.8kb XbaI fragment and pANC4 the 2.3kb BamHI - XbaI fragment of pANC1 (see Figure 3.1.4). Transformants of strain C43 obtained with both pANC3 and pANC4 showed complementation of creA204 but not argB2. The insert of pANC4 was used for further analyses since the small 0.5kb BamHI - XbaI fragment in pANC3 represented a sequence which would not have been continuous with the 2.3kb BamHI - XbaI fragment in the genome, having resulted from circularization at the BamHI site when the plasmid was generated. This was confirmed by comparing Southern blots of wildtype A.
Figure 3.1.5: Southern blot analysis of wildtype *A. nidulans* and transformant T41. Each track contains 2ug of DNA digested with *XbaI*. The probe used was pANCl.

Tracks:
1. T41
2. wt *A. nidulans*
3. C43

Bacteriophage lambda DNA digested with *HindIII* was used as a marker of molecular weight.
*nidulans* DNA digested with various restriction enzymes and probed with pANC3 and pANC4. Faint bands of hybridization to wildtype DNA are seen when probed with pANC3 which are not present when probed with pANC4 (Figures 3.1.7 and 4.1.1). A partial restriction map of the pANC4 insert is shown in Figure 3.1.6. These analyses also demonstrate that all bands can be accounted for by the restriction map of pANC4 and therefore it was unlikely that the insert had rearranged during the rescue procedures.

pANC4 was also found to complement the creA220, creA225, creA<sup>d</sup>-1 and creA<sup>d</sup>-30 alleles. This ruled out the possibility that it contained an allele specific suppressor of creA. Hybridization of pANC4 to genomic XbaI digests of DNA from strains carrying the alleles creA220, cre225 and creA<sup>d</sup>-1 (Figure 3.1.8) showed that these mutants were not the result of any gross deletions, insertions or rearrangements in the genome since hybridization was seen to the same 10.5kb band as seen for the wildtype and the creA204 strains.

3.2 ANALYSIS OF CreA<sup>+</sup> TRANSFORMANTS:

DNA was isolated from 16 phenotypically CreA<sup>+</sup> pANC3 transformants of strain C43, designated TC43A - TC43P. Slot blot analysis of serial dilutions of the DNA samples using the insert of pANC4 as a probe showed that all transformants contained transforming sequences and the approximate number of copies of creA in these transformants was determined. DNA loadings were standardized using the single copy *ribob* gene of *A. nidulans* from pPL3. Copy numbers ranged from 1 for those that were the result of a gene replacement event to about 15 in the case of TC43E (Table 3.2.1).
Figure 3.1.6: Partial restriction map of the insert in pANC4. The direction of transcription of the creA gene is indicated by the arrow.

Restriction site abbreviations:

- B - BamHI
- E - EcoRV
- Ps - PstI
- Pv - PvuII
- R - Rsal
- S - SalI
- X - XbaI

The XbaI site is shown in brackets since it is not a genomic site and resulted from the subcloning of pANCl.
Figure 3.1.7: Southern blot analysis of wildtype *A. nidulans* DNA cut with various restriction enzymes and probed with pANC3. Each track contains 2ug of DNA cut with:

1. *Bam*HI
2. *Bgl*II
3. *Eco*RI
4. *Hind*III
5. *Pst*I
6. *Pvu*II
7. *Sac*I
8. *Sma*I
9. *Sal*I
10. *Xba*I
11. *Xho*I
12. *Cla*I

Bacteriophage lambda DNA digested with *Hind*III was used as a marker of molecular weight.
Figure 3.1.8: Southern blot analysis of XbaI digested DNA from wildtype A. nidulans and strains containing creA mutant alleles. The amounts of DNA in each track are not equal.

Tracks are:

1-3. wildtype A. nidulans.

4. 664 (creA204).

5. 1070 (creA^d-1).

6. C43 (creA204).

7. creA204.

8. creA220.

9. creA225.

Bacteriophage lambda DNA digested with HindIII was used as a marker of molecular weight.
Table 3.2.1: creA copy number in CreA⁺ transformants.

<table>
<thead>
<tr>
<th>COPY NUMBER</th>
<th>TRANSFORMANTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TC43A, TC43J, TC43M.</td>
</tr>
<tr>
<td>2</td>
<td>TC43B, TC43F, TC43K.</td>
</tr>
<tr>
<td>3-6</td>
<td>TC43C, TC43D, TC43G, TC43H, TC43I, TC43L, TC43N,</td>
</tr>
<tr>
<td></td>
<td>TC43O, TC43P.</td>
</tr>
<tr>
<td>15</td>
<td>TC43E.</td>
</tr>
</tbody>
</table>
Southern hybridization of XbaI digested DNA using pANC4 as a probe showed all transformants contained pANC4 sequences and retained the 10.5kb native creA band except TC43H, in which the native band was disrupted by transforming DNA (Figure 3.2.1). TC43H was outcrossed to strain J7 and among 485 progeny scored, none were of creA mutant phenotype. This indicated that this transformant arose from a homologous integration at the creA locus, and that the clone was of creA and not a suppressor of the creA mutant phenotype. TC43H was meiotically stable, as no progeny of creA mutant phenotype were found among 465 progeny of a selfed cleistothecia. These results, where integration has occurred at the homologous genomic sequence, and the complementing sequence is inseparable from the creA mutation through meiosis is further evidence that the clone contains creA.

75 pANC3 transformants of the creA204 strain were tested for their ability to utilize a range of compounds as sole carbon sources at a concentration of 1% in appropriately supplemented minimal media with 10mM ammonium tartrate as the nitrogen source. Compounds tested were D-glucose, glycerol, D-sorbitol, sucrose, ethanol, maltose, acetate, lactose (at 0.5%), D-mannitol, D-galactose, D-melibiose, raffinose, L-arabinose, L-sorbose and quinic acid. Utilization of L-threonine, acetamide, 2-pyrrolidinone, B-alanine, p-aminobenzoic acid, L-proline and L-glutamic acid were tested as both carbon and carbon and nitrogen sources at a final concentration of 50mM in the media. There was variation between transformants on synthetic complete medium, and only these same differences were evident for all the carbon sources tested. Therefore, none of the creA+ transformants could be identified as showing appreciably different utilization of these carbon sources as compared to wildtype A.
Figure 3.2.1: Southern blot analysis of *A. nidulans* pANC4 transformants. Each track contains 2ug of DNA digested with *XbaI*. The probe used was pANC4.

Tracks are:

1. wildtype *A. nidulans*
2. C43
3. T41
4. TC43A
5. TC43B
6. TC43C
7. TC43D
8. TC43E
9. TC43F
10. TC43G
11. TC43H

Bacteriophage lambda DNA digested with *HindIII* was used as a marker of molecular weight.
nidulans. An increase in the number of copies of creA may not necessarily be expected to lead to detectable differences in the utilization of various carbon sources since the creA mutant alleles themselves do not show any such differences in growth (Arst and Cove 1973, Bailey and Arst 1975, Hynes and Kelly 1977). This homogeneous phenotype was also to be expected if the creA gene product has a regulatory role as opposed to being directly involved in the metabolism of carbon sources. Growth differences in plate testing may therefore only be detected by other approaches such as titration analyses in multicopy transformants or by specific plate testing which may show variation in the extent to which metabolic functions are subject to carbon catabolite repression by creA. Transformants TC43A - TC43H were tested for their ability to grow on sucrose media containing allyl alcohol at concentrations ranging from 2.5mM to 100mM. Those transformants with more than two extra copies of creA grew consistently better than wildtype on these media, suggesting tighter repression of alcohol dehydrogenase synthesis.

3.3 EXPRESSION OF creA IN A. nidulans:

Strand specific probes were generated from the insert of pANC4 which had been cloned into M13 mp18 and mp19. Each of these was used separately as probes to slot blots of total RNA isolated from wildtype A. nidulans grown with glucose as the sole carbon source. The differential hybridization between these probes to RNA showed that the creA gene is transcribed in the direction from the BamHI site to the XbaI site, with respect to the restriction map of pANC4 (see Figure 3.1.6).
Total RNA was isolated from wildtype and creA mutant strains grown in the presence of various carbon sources with 10mM ammonium tartrate or 10mM ammonium orthophosphate for acetate medium acting as the nitrogen sources. Northern blot analysis using the insert of pANC4 as a probe showed that wildtype A. nidulans and the mutant strain creA204 produce a major mRNA species of approximately 2.0kb in length (Figure 3.3.1). Under the growth conditions used, creA mRNA levels in wildtype A. nidulans in the presence of glucose were of the same order of magnitude as that for the argB transcript, in that similar signals were given by hybridization to both transcripts when probed with pANC1 (Figure 3.3.2). Relative levels of creA message were determined by slot blot analysis of RNA samples probed with the insert of pANC4. Levels were calculated relative to wildtype A. nidulans RNA from 1% D-glucose grown mycelia (Table 3.3.1). Two fold higher levels were found in 1% L-arabinose and 1% glycerol grown cultures, but not in 50mM D-quinate or 50mM L-proline grown cultures. Similar levels of creA expression were found for the mutant strain creA204 as for wildtype A. nidulans in D-glucose grown mycelia. In glycerol, L-arabinose and acetate grown cultures of creA204, creA expression was increased 1.5 - 2.5 fold as in the wildtype strain. The altered levels of the creA message present under these different growth conditions as determined by Northern analyses are insufficient to account for the regulatory actions of creA and its phenotypic role in carbon catabolite repression. Although transcriptional regulation is still possible, the activation/inactivation of the CreA protein in response to some external stimulus is likely to be involved in the regulation of carbon catabolite repression by creA.

Slot blot and Northern analyses were also performed using the
Figure 3.3.1: Northern blot analysis of RNA isolated from wildtype 
A. nidulans and creA204 mutant strain. Each track contains 
approximately 20ug total RNA. The probe used was the insert of 
pANc4.

Tracks are:

1. wildtype A. nidulans from D-glucose grown mycelia.
2. wildtype A. nidulans from glycerol grown mycelia.
3. wildtype A. nidulans from L-arabinose grown mycelia.
4. creA204 from D-glucose grown mycelia.
5. creA204 from glycerol grown mycelia.
6. creA204 from L-arabinose grown mycelia.
Figure 3.3.2: Northern blot analysis of RNA isolated from wildtype *A. nidulans* D-glucose grown mycelia. The probe used was pANCl. The *creA* and *argB* transcripts are indicated.
Table 3.3.1: Relative levels of expression of creA in wildtype *A. nidulans* and strain 664.

<table>
<thead>
<tr>
<th>CARBON SOURCE</th>
<th>WT.</th>
<th>664</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-glucose</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>L-arabinose</td>
<td>2.5</td>
<td>1.5</td>
</tr>
<tr>
<td>glycerol</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>acetate</td>
<td>2.5</td>
<td>1.5</td>
</tr>
<tr>
<td>L-proline</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>quinic acid</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Relative values are rounded to the nearest 0.5 of a unit.

Table 3.3.2: Relative levels of expression of creA in creA mutant strains and CreA+ transformants.

<table>
<thead>
<tr>
<th>CARBON SOURCE</th>
<th>1070</th>
<th>creA220</th>
<th>creA225</th>
<th>TC43D</th>
<th>TC43E</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-glucose</td>
<td>1.0</td>
<td>---</td>
<td>3.0</td>
<td>20.0</td>
<td>6.0</td>
</tr>
<tr>
<td>L-arabinose</td>
<td>1.0</td>
<td>4.0</td>
<td>1.5</td>
<td>---</td>
<td>30.0</td>
</tr>
</tbody>
</table>

Relative values are rounded to the nearest 0.5 of a unit.
insert of pANCl4 as a probe to RNA isolated from the strains carrying the mutant alleles creA\textsuperscript{d-1}, creA220 and creA225 (Table 3.3.2 and Figure 3.3.3). Strain 1070, carrying the creA\textsuperscript{d-1} allele, had levels of creA mRNA comparable with D-glucose grown wildtype \textit{A. nidulans} when both D-glucose and L-arabinose were present as carbon sources. However, in the presence of D-glucose, but not L-arabinose, 1070 expresses another transcript which is larger than the native creA message. A larger message was also present in the strains carrying the creA220 and creA225 alleles which is expressed regardless of whether D-glucose or L-arabinose acts as the carbon source. creA225, in contrast to strains carrying the other mutant alleles, showed higher levels of creA expression in the presence of D-glucose than when L-arabinose was the sole carbon source. These analyses showed that, of the mutant alleles available, all expressed creA transcript(s) in some form. It is therefore possible that these alleles may not represent complete loss of function alleles and that the derepressed phenotypes of the creA mutants may not be the result of a true null allele.

Total RNA was also prepared from the creA multicopy transformants TC43E and TC43D to determine whether or not increased copy number led to increased levels of creA messenger RNA. TC43E and TC43D grown with D-glucose as the carbon source had approximately six times, and twenty times respectively, the level of creA mRNA expression as that found in similarly grown wildtype \textit{A. nidulans}. This was increased to thirty times that found in wildtype D-glucose grown mycelia when TC43E was grown with L-arabinose as the carbon source (Table 3.3.2). Thus despite showing increased levels of creA expression TC43E still showed regulated creA expression. The level of
Figure 3.3.3: Northern blot analysis of RNA isolated from wildtype *A. nidulans* and creA mutant strains probed with the insert of pANC4. Each track contains approximately 20ug of total RNA.

Tracks are:

1. Wildtype *A. nidulans* from D-glucose grown mycelia.
2. creA220 from D-glucose grown mycelia.
3. creA220 from L-arabinose grown mycelia.
4. creA225 from D-glucose grown mycelia.
5. creA225 from L-arabinose grown mycelia.
6. creAd-1 from D-glucose grown mycelia.
7. creAd-1 from L-arabinose grown mycelia.
creA expression found in TC43D, with two copies compared to 15 in TC43E, showed that higher message levels did not necessarily correlate with higher creA copy number or with the observation of apparently "tighter" repression on sucrose plus allyl alcohol media. This, along with the observation that TC43D and TC43E produce multiple transcripts are probably both effects of the positions of integration of transforming sequences (Figure 3.3.4).

3.4 SUMMARY AND CONCLUSIONS:

The creA gene from Aspergillus nidulans has been cloned by complementation of a non-revertable mutant allele using a genomic library and plasmid rescue techniques. The rescued sequence was subcloned and a 2.3kb fragment was identified which complements several creA mutant alleles. Northern analyses showed that creA encodes a transcript of approximately 2.0kb in length and that the levels of this transcript varied by up to two fold, depending on the carbon source. All the mutant alleles that were tested showed expression of a creA transcript. Transformants containing one or more copies of creA grew as wildtype on a large range of carbon sources, but for transformants with two or more extra copies of creA there was evidence for tighter carbon catabolite repression.
Figure 3.3.4: Northern blot analysis of RNA isolated from wildtype *A. nidulans* and pANC4 transformants of strain C43. Tracks contain 20μg of total RNA. The insert of pANC4 was used as a probe.

Tracks are:

1. Wildtype *A. nidulans* from D-glucose grown mycelia.
2. TC43D from L-arabinose grown mycelia.
3. TC43D from D-glucose grown mycelia.
4. TC43E from D-glucose grown mycelia.
5. TC43E from L-arabinose grown mycelia.
CHAPTER FOUR
ANALYSIS OF creA<sup>d</sup>-30 AND THE
PHENOTYPE OF A creA NULL ALLELE.

Experiments using the creA<sup>d</sup>-30 allele are presented in this chapter and they provide additional evidence that the clone isolated and described in Chapter 3 is of the creA region. The molecular analysis of the mutant allele creA<sup>d</sup>-30 along with the findings in the previous chapter that all mutant alleles express creA transcripts, demonstrates the importance of constructing a true loss of function creA allele. Work presented in this section, where a loss of function creA allele was constructed by reverse genetics, shows that creA is an essential gene, and that homologous genes are probably present in a number of other fungi, especially those more closely related to A. nidulans.

4.1 MOLECULAR LOCALIZATION OF THE INVERSION BREAKPOINT IN creA<sup>d</sup>-30:

The creA<sup>d</sup>-30 mutation is the most phenotypically extreme creA allele known and results in a very compact colony morphology. creA<sup>d</sup>-30 was identified by Arst and co-workers (Arst et al. 1990) as a spontaneous mutation in a strain containing the frA-1 mutation. frA-1 leads to the inhibition of growth by D-fructose, D-sorbitol, sucrose and D-mannitol (Roberts 1963). The creA<sup>d</sup>-30 mutation conferred resistance to the toxicity of 1% D-mannitol in the presence of 1% L-
arabinose with 5mM L-glutamate as the nitrogen source in a frA-1 strain. The basis for this resistance is unknown. Linkage analysis showed that the creA<sup>d</sup>-30 mutation is a result of a pericentric inversion, with one breakpoint occurring in creA on the left arm of chromosome I and the other occurring between the markers binG and yA on the right arm of chromosome I. Since creA<sup>d</sup>-30 was the result of an inversion with one breakpoint occurring within the creA locus it has been presumed that the phenotype, of derepression, is that of a loss of function allele. The identification and genetic characterization of creA<sup>d</sup>-30 in Arst's laboratory provided a means to show unequivocally that the cloned fragment was a clone of the creA region. Using the cloned gene, the inversion breakpoint of creA<sup>d</sup>-30 could be mapped to a particular position within the creA gene itself. The clone could then be used to determine whether the creA<sup>d</sup>-30 allele was a true loss of function allele at the level of mRNA, an important point underpinning the genetical hypothesis of negative control.

Genomic DNAs from wildtype A. nidulans and creA<sup>d</sup>-30 were singularly and doubly digested with a number of restriction enzymes, sites for which were present within the insert of pANC4, and Southernns of these digests were probed with the insert of pANC4 (Figure 4.1.1). For each of the single and double digests, the patterns of hybridization to the insert of pANC4 differed between the two strains, indicating that the cloned fragment represented a region which was disrupted in creA<sup>d</sup>-30. The expected patterns of hybridization to digests of wildtype DNA, with respect to the restriction map of pANC4 are shown in Table 4.1.1, and all bands were present in the wildtype strain. The altered bands found in the mutant strain creA<sup>d</sup>-30 are also shown. These altered bands represent fusion
Figure 4.1.1: Southern blots of wildtype A. nidulans and creA\textsuperscript{d-30} DNA probed with pANC4. For each pair of tracks DNA from the wildtype creA strain is on the left and from the creA\textsuperscript{d-30} strain on the right. Digests are:


In the longer exposure tracks of PstI single and double digests, faint bands of hybridization to a 150bp fragment in the creA\textsuperscript{d-30} strain are visible, while in the wildtype a doublet (150bp and 250bp) is present). In interpreting these blots several other points should be noted: 1. Tracks B3 and B8 wildtype: the faint band at 4.4kb results from partial digestion and is not present in other blots. 2. Tracks B3, B8, C5, C8 and C9 creA\textsuperscript{d-30}: the faint band below the strong 1kb band is a fusion band. 3. Track B7 creA\textsuperscript{d-30}: two fusion bands are present - a faint one at approximately 1kb and a strong one forming a doublet with the largest band. 4. Track B4 creA\textsuperscript{d-30}: the "band" at 4.4kb is not present in other blots and is believed to be an artifact whilst the strong band is probably a doublet containing a fusion band. Only one fusion band is apparent in PstI and PvuII digests of creA\textsuperscript{d-30} DNA, because the breakpoint is probably too close to the PvuII/PstI/PvuII site cluster to allow hybridization to the other fusion band.
Table 4.1.1 Expected bands of hybridization of pANC4 (insert) to wildtype *A. nidulans* DNA. For each restriction enzyme the sizes of the expected bands of hybridization to *A. nidulans* genomic DNA are listed. These sizes were generated from the restriction map of pANC4. *indicates the bands altered in creA<sup>d</sup>-30 digests.

<table>
<thead>
<tr>
<th>RESTRICTION ENZYME(S)</th>
<th>BANDS OF HYBRIDIZATION (KB)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>BamHI</em></td>
<td>&gt;2.3*</td>
</tr>
<tr>
<td><em>EcoRV</em></td>
<td>&gt;1.5* &gt;0.425 0.475</td>
</tr>
<tr>
<td><em>PstI</em></td>
<td>&gt;0.8 0.15 0.25* &gt;1.1</td>
</tr>
<tr>
<td><em>PvuII</em></td>
<td>&gt;1.1 &gt;1.2 (band in mutant is a doublet)</td>
</tr>
<tr>
<td><em>XbaI</em></td>
<td>&gt;2.3*</td>
</tr>
<tr>
<td><em>Xhol</em></td>
<td>&gt;2.3*</td>
</tr>
<tr>
<td><em>BamHI/EcoRV</em></td>
<td>1.5* 0.475 &gt;0.425</td>
</tr>
<tr>
<td><em>BamHI/PstI</em></td>
<td>0.8 0.15 0.25* &gt;1.1</td>
</tr>
<tr>
<td><em>BamHI/PvuII</em></td>
<td>1.2* &gt;1.1</td>
</tr>
<tr>
<td><em>BamHI/XbaI</em></td>
<td>&gt;2.3*</td>
</tr>
<tr>
<td><em>EcoRV/XbaI</em></td>
<td>&gt;1.5* 0.475 &gt;0.425</td>
</tr>
<tr>
<td><em>PvuII/PstI</em></td>
<td>&gt;0.8 0.15 0.25* &gt;1.1</td>
</tr>
<tr>
<td><em>PstI/XbaI</em></td>
<td>&gt;0.8 0.15 0.25* &gt;1.1</td>
</tr>
<tr>
<td><em>PvuII/XbaI</em></td>
<td>&gt;1.2 &gt;1.1</td>
</tr>
</tbody>
</table>
fragments between the creA cloned region and stretches of DNA flanking the creA\textsuperscript{d}-30 inversion breakpoints. It was evident that in the \textit{PstI} single and double digests the creA\textsuperscript{d}-30 strain only showed hybridization to the 150bp fragment of the two smaller bands, while the wildtype strain showed hybridization to the same 150bp fragment plus a 250bp fragment. Bands of hybridization present in all other digests could be accounted for by a breakpoint occurring in the 250bp \textit{PstI} fragment of the cloned gene. Since only one fusion band is apparent in the \textit{PvuII} and \textit{PstI} digests of creA\textsuperscript{d}-30 DNA, the breakpoint within the 250bp \textit{PstI} fragment is probably located very close to the \textit{PvuII/PstI/PvuII} (overlapping sites identified by sequencing) cluster and as such hybridization to one of the fusion bands could not be detected.

Total RNA was prepared from wildtype \textit{A. nidulans} and the creA\textsuperscript{d}-30 strain grown with 1\% D-glucose and 1\% L-arabinose serving as the carbon sources in minimal medium. Northern analysis using the \textit{pANC4} insert as a probe (Figure 4.1.2) showed that creA message was present in this strain, but the creA\textsuperscript{d}-30 mutation gives rise to an altered creA transcript. When cultures were grown with 1\% L-arabinose as a carbon source the creA\textsuperscript{d}-30 strain expressed a transcript which is larger than the wildtype creA message. When D-glucose is used as the carbon source this message is also present together with an additional transcript which is smaller than the wildtype message. The presence of these two transcripts in D-glucose grown cultures may indicate that the 3'end of creA has been fused to a gene which is expressed when D-glucose but not L-arabinose is present as the carbon source. The larger message, which is expressed in both growth conditions, probably represents the 5' end of the creA transcript.
Figure 4.1.2: Northern blot analysis of wildtype *A. nidulans* and creA<sup>d</sup>-30 RNA probed with the insert of pANC4.

Tracks are:

1. Wildtype *A. nidulans* from D-glucose grown mycelia.
2. creA<sup>d</sup>-30 from D-glucose grown mycelia.
3. Wildtype *A. nidulans* from L-arabinose grown mycelia.
4. creA<sup>d</sup>-30 from L-arabinose grown mycelia.
which has extended through the inversion breakpoint in creA\textsuperscript{d}-30. Therefore, the creA\textsuperscript{d}-30 mutation does not represent a formal lack of function allele making it necessary to construct a gene replacement strain to determine the phenotype of a lack of function mutation. Chromosomal rearrangements leading to the expression of truncated protein products do not necessarily represent loss of function alleles and this has been demonstrated for a number of mutant genes. For example, the areA\textsuperscript{r}-18 mutation results from a translocation breakpoint within the areA locus and leads to an inability to utilize nitrogen sources other than ammonium and glutamine. areA\textsuperscript{r}-18 can revert intracistronically via further rearrangements involving the 3'end of the gene such that the gene is transcribed and a functional protein expressed (Arst \textit{et al.} 1989) which lacks the non-essential 336 N-terminal amino acids (Kudla \textit{et al.} 1990). Studies have also shown that a large deletion of the 3' end of the alcR gene resulted in a gene still capable of weak partial complementation of the alcR125 allele (Lockington 1984). Furthermore, many oncogenes are cellular genes that have been activated by translocation (reviewed by Haluska \textit{et al.} 1987).

4.2 CONSTRUCTION OF A creA DELETION STRAIN:

4.2.1 The isolation of a larger creA genomic clone:

The creA clone isolated by marker rescue was only 2.3kb in size. Although this clone contained the required sequences to complement the creA mutant alleles, Northern analyses showed that this genomic clone was not much larger than the corresponding message it detected. It was therefore necessary for the construction of a creA deletion
strain to isolate a larger genomic clone which contained a sufficient amount of creA flanking sequence.

The creA clone in pANC4 was shown to hybridize to a 7.5kb EcoRI fragment by Southern analysis. DNA from wildtype *A. nidulans* was digested with EcoRI to completion and fragments in the size range 6 – 8kb were purified. These fragments were cloned into the EcoRI site of pUC19 and transformed into *E. coli* strain JM101 to create a partial plasmid library. 900 colonies harbouring recombinant plasmids were screened using the insert of pANC4 as a probe. 4 colonies were identified by this method as harbouring sequences homologous to the pANC4 insert. Southern analysis of mini plasmid preparations from these colonies showed that two of these contained EcoRI inserts of the correct size which were homologous to the original clone in pANC4. This plasmid was designated pANC6 and a partial restriction map is shown in Figure 4.2.1.

4.2.2 The construction of a creA gene replacement vector:

The *ribB* gene in pPL3 was chosen as the transformation selection marker to be used for the construction of a plasmid to replace the wildtype *A. nidulans* creA gene. The 7.5kb EcoRI insert of pANC6 was purified and ligated into the EcoRI site of pUCBX (pUC19 in which the *BamHI* and *XbaI* sites had been deleted from the polylinker). This plasmid was designated pANC7 and was digested with *BamHI* and *XbaI* to cut out the 3.5kb *BamHI-XbaI* fragment containing the creA gene. The larger fragment generated, containing pUCBX along with sequences flanking the creA gene, was purified and end filled. This was then blunt end ligated with the 2.3kb *KpnI* fragment of pPL3 and
Figure 4.2.1: Partial restriction map of the insert in pANC6. Triangles indicate the extent of the insert in pANC4.

Restriction site abbreviations:

- B - BamHI
- E - EcoRI
- Ev - EcoRV
- H - HindIII
- Ps - PstI
- Pv - Pvull
- S - SalI
- Sc - SacI
containing the riboB gene which had also been end filled. The ligation mixture was transformed into *E. coli* strain DH1 and colonies harbouring the desired plasmid construct were identified using duplicate colony blots probed separately with the 2.3kb *KpnI* insert of pPL3 and the insert of pANC6. A restriction map of this 9.2kb plasmid, pANC8, is shown in Figure 4.2.2. It contains the riboB gene inserted between 1.95kb and 2.1kb of sequence normally flanking the creA gene in the genome. pANC8 was also used as a probe to Southernns of pUC19, pANC6, pANC7 and pPL3 digested with the appropriate restriction enzymes in order to confirm that the desired construct was correct. pANC8 used as a probe to Northernns of wildtype *A. nidulans* total RNA showed that only one message was detected which corresponded to the riboB transcript (Figure 4.2.3). Therefore, pANC8 did not contain any sequences which, if transcribed, could produce message homologous to creA.

4.2.3 The creation of a creA gene replacement strain of *A. nidulans*:

The *A. nidulans* strains C61 (*yAl; pyroA4; nicB8; niiA4 riboB2) and C62 (*biAl; AcrAl; galE1; facA303; sb3; riboB2) were both derived from a cross between J7 and MSF and were transformed with the gene replacement construct pANC8. RiboB+ transformants were selected on appropriately supplemented protoplast media. Thirty RiboB+ transformants of strain C61 (TC611 - TC6130) and twenty eight of strain C62 (TC621 - TC6228) were purified for further analysis. On minimal medium all of these transformants except TC622 showed wildtype morphology. The colony morphology of TC622 was extremely compact, not unlike that expected for a creA mutant strain. Similarly, this transformant showed very weak growth on sucrose
Figure 4.2.2: Partial restriction map of the insert in pANC8. The open region contains the riboB gene of *A. nidulans* and the coloured regions are sequences flanking the creA gene in pANC6.

Restriction site abbreviations:

- **B** - *BamHI*
- **E** - *EcoRI*
- **Ev** - *EcoRV*
- **H** - *HindIII*
- **Ps** - *PstI*
- **Pv** - *PvuII*
- **S** - *SalI*
- **Sc** - *SacI*
Figure 4.2.3: Northern blots of wildtype *A. nidulans* RNA isolated from D-glucose grown mycelia. Each pair of tracks contains approximately 20ug of total RNA.

Panels were probed with:

A. *KpnI* fragment from pPL3 containing the *riboB* gene.
B. pANC8.
C. *BamHI-XbaI* fragment from pANC6 containing the *creA* gene.
medium containing 2.5mM allyl alcohol, while the rest of the transformants had wildtype levels of growth on this medium. DNA was isolated from each of these 58 RiboB\(^+\) transformants. Southern's of DNA samples digested with EcoRI were probed with the 3.5kb BamHI - XbaI fragment from pANC7 in order to determine whether or not any of these transformants were the result of a gene replacement event at the cre\(A\) locus. All the transformants, including TC622, showed hybridization to the native cre\(A\) EcoRI fragment at about 7.5kb and therefore no transformant had the native cre\(A\) gene replaced with riboB (for example see Figure 4.2.4). Transformants resulting from a gene replacement event with pANC8 would not be expected to show hybridization to this BamHI-XbaI fragment from pANC7 (Figure 4.2.5).

At the same time as these experiments were being undertaken, pANC8 was also used to transform a diploid strain since the possibility existed that a true cre\(A\) null allele may be lethal, in which case a gene replacement event could not be obtained in a haploid strain.

A diploid strain, C63, was constructed between strains C61 and C62 and used in transformation experiments with pANC8. RiboB\(^+\) transformants of C63 were selected on protoplast minimal medium with 10mM sodium nitrate as the nitrogen source, thus maintaining selection for the diploid state. Fifty of these transformants (TC631 - TC6350) were purified and picked onto 1% complete medium containing various concentrations of benlate for the formation of haploid sectors. The rationale for this was that if a cre\(A\) gene replacement event had occurred in one of the chromosomes(I) in the diploid, and this event was lethal in haploid cells, then transformants which were the result of a gene replacement could be identified as segregating either only green or only yellow haploid
Figure 4.2.4: Southern blot of EcoRI digested DNA from pANC8 transformants of strain C62 probed with the BamHI-XbaI fragment of pANC6 containing the creA gene. The amounts of DNA in each track are not equal.

Tracks are:

1. C62
2. TC6212
3. TC6213
4. TC6214
5. TC6215
6. TC6216
7. TC6217
8. TC6218
9. TC6219
10. TC6220
11. TC6221
12. TC6222

Bacteriophage lambda DNA digested with HindIII was used as a marker of molecular weight.
Figure 4.2.5: Strategy used for deleting the creA gene from *A. nidulans*.

(a). The *creA* gene present within a 7.5kb *EcoRI* fragment in the genome.

(b). Insert of pANC8 constructed to replace the *creA* gene. It contains the *ribob* gene inserted between sequences that flank *creA* in the genome.

(c). The 2.6kb and 3.3kb *EcoRI* fragments expected in the genome of a *creA* deletion strain created by recombination events between flanking sequences in (a) and (b).
sectors on complete benlate medium. Since conidial colour distinguishes haploids on the basis of which chromosome I is present, transformants could be screened by eye.

Of the 50 RiboB⁺ transformants screened, two failed to produce haploids of both colours. TC6330 fails to sector green while TC6332 fails to sector yellow haploids. Both these transformants segregated all other chromosomes as determined from the genotypes of haploid progeny. All haploids derived from both transformants were phenotypically Ribo⁻ and therefore the transforming sequences were located on the chromosome I not represented in haploids.

Similarly, TC6330 and TC6332 also failed to produce haploids of both colours when haploids were selected on benlate medium containing glycerol or L-arabinose or fructose as the sole carbon source, eliminating the possibility that haploids unable to utilize particular carbon sources were segregating. The failure to segregate haploids of both colours was determined not to be temperature specific by the selection of haploids at 25°C, 30°C, 37°C and 42°C on media containing the various carbon sources.

DNA was isolated from strains C61, C62 and C63 and the transformants TC6330 and TC6332. Duplicate Southernss of EcoRI digests of these samples were probed with (I) pMD06 plus the 3.5kb BamHI - XbaI fragment of pANC7 and (II) pUC19 plus the 3.5kb BamHI - XbaI fragment of pANC7 (Figure 4.2.6). The autoradiogram of filter (I) showed that approximately equal amounts of each DNA sample were present as shown by the intensities of the larger band corresponding to the single copy of the argB gene. Hybridization to the creA native band on this filter showed that, at least in the case of TC6330,
Figure 4.2.6: Southern analyses of gene disrupted diploids TC6330 and TC6332. Each track contains 2ug of DNA digested with EcoRI from:

1. C63
2. TC6330
3. TC6332
4. C61
5. C62

From right to left the panels were probed with:

1. pMO06 and the 3.5kb BamHI-XbaI fragment from pANC7 containing the creA gene.
2. pUC19 and the 3.5kb BamHI-XbaI fragment from pANC7 containing the creA gene.
3. The 6.9kb BamHI-XbaI fragment of pANC7 containing pUC19 and sequences flanking creA in the genome
4. pUC19.

N.B. While the band of hybridization corresponding to the native creA band in C61 is not as intense as that for C62 in panels 1 and 2, equal intensities were consistently seen on repeat filters.

Autoradiograms were exposed without the use of intensifying screens.
there appeared to be only half the amount of hybridization when compared to the diploid parent strain. The smaller band hybridizing in TC6332 represents pUC19 sequences as shown when the same samples were probed with pUC19 alone. TC6330 did not contain any pUC19 sequences and therefore appeared to be the result of a direct gene replacement event of one of the native creA genes in the diploid. Further to this, when the same samples were probed with the larger BamHI - XbaI fragment of pANC7 (containing pUCBX and the creA flanking regions) TC6330 showed about half the amount of hybridization to the native band at 7.5kb as compared to the parent strain and to two smaller bands of 2.6kb and 3.3kb. It can be determined from the restriction map of the insert of pANC8 that this is the expected pattern of hybridization for a gene replacement diploid strain (Figure 4.2.5). TC6332 probably has more than one copy of pANC8. While TC6330 showed patterns of hybridization expected for a gene replacement event, it may also have a small amount of creA flanking sequences integrated into another site in the genome. Autoradiograms exposed for more than a week showed a very faint band of hybridization at approximately 500bp, to TC6330 DNA probed with creA flanking sequences. DNA from six haploids derived from TC6330, each having a unique chromosome complement were screened by Southern analysis using the creA flanking sequences as a probe to determine to which linkage group these extra sequences could be assigned. None of the haploids contained extra transforming sequences and therefore these must have been integrated into the same chromosome I as the gene replacement event in the diploid. As far as can be determined this does not interfere with the gene replacement event. Therefore, analysis of TC6330 at the molecular level demonstrated that it is a transformant resulting from a gene replacement event of one of the
The gene disruption plasmid was used to transform two haploid strains. Of the many transformants obtained, none had a creA mutant phenotype or a true disruption of the CreA locus. This result suggests the hypothesis that a disruption of the CreA locus tp give a null allele results in a lethal phenotype. This hypothesis would predict that a gene disruption could be obtained in a diploid where the undisrupted allele could complement the null allele. Such a diploid would have the phenotype that it could not be broken down to give a haploid sector bearing the chromosome carrying the disrupted allele. This experiment was carried out and two diploid transformants with this phenotype were obtained. As predicted by the hypothesis, both of these transformants were shown by Southern analysis to have a gene replacement of one of the two creA alleles. Thus it is likely that the creA product is essential for cell viability.

The creA mutant alleles isolated by various selection procedures would therefore represent a class of alleles with altered function rather than total loss of creA function for which mutants could not have been identified in the haploid state. This raises the question as to whether creA acts in a positive or negative manner in regulating the expression of genes subject to carbon catabolite repression. Previous analyses have assumed creA alleles to result from loss of function.

4.2.4 Vegetative versus conidial lethality of the creA null phenotype:

Although it was quite evident that a creA null allele could not
be segregated out in haploidization analysis, the question remained as to whether the phenotype was lethality or failure to form conidia. It may, in some instances, be difficult to distinguish between haploid aconidial mutants and aneuploid sectors during haploidization. However, when non-conidiating sectors were picked onto 1% complete medium they failed to show the vegetative mycelial growth expected for aconidial mutants and were probably aneuploid. In addition to this, a large number of non-conidiating sectors from haploidization plates were scraped into a suspension and plated out at low density onto 1% complete medium with no evidence of colonies which grew but failed to conidiate.

A further experiment to show that the creA null allele is lethal in vegetative cells prior to the development of conidia was carried out using mycelial cultures of TC6332. TC6332 was used in this series of experiments in preference to TC6330 so that haploids containing the deletion allele could be selected for by the absence of both riboflavin and biotin from the growth media. Haploids from this diploid containing the wildtype allele, which is on the same chromosome as the biAl marker, are auxotrophic for biotin and are selected against under these conditions. TC6332 was grown overnight in liquid cultures supplemented with sodium thiosulphate, pyridoxal hydrochloride, nicotinic acid and ammonium tartrate as the nitrogen source and various concentrations of benlate (or without benlate), thereby allowing haploid cells with the genotype biA⁺ riboB⁺ to grow simultaneously selecting against biAl haploids having the wildtype creA allele. Mycelial cultures were harvested, ground coarsely in a coffee grinder and spread onto 1% similarly supplemented minimal medium containing various concentrations of benlate. Mycelia that
continued to grow on such media must therefore be biA\(^+\) and riboB\(^+\). All of the resulting mycelial growth on these plates eventually conidiated and were found to be diploid. No biA\(^+\) riboB\(^+\) areas of growth were haploid or failed to conidiate. This was consistent with the hypothesis that a creA null allele is lethal in vegetative cells as opposed to failing to form conidial progenitor cells or conidia themselves. If the latter was the case then areas of haploid vegetative growth should be identifiable. These experiments suggest therefore that the creA null allele is recessive lethal in vegetative cells.

4.2.5 Rescuing the ability of TC6330 to form haploids containing chromosome I having the gene replacement:

The failure of TC6330 to form haploids containing one member of the chromosome I pair was due to the inability to form haploids containing a creA gene replacement. Such a situation should be able to be rescued by inserting a wildtype copy of creA into any chromosome except the chromosome I carrying the wildtype allele in TC6330.

In an attempt to demonstrate this, TC6330 was co-transformed with the wildtype creA gene in pANC7 and pAmPh366-1 conferring resistance to blenoxane as a dominant selectable marker. Diploid transformants were selected on minimal protoplast medium containing sodium nitrate as the nitrogen source and blenoxane. Twenty bleomycin resistant transformants were picked onto 1% complete medium containing benlate. Four of these diploid transformants appeared to segregate both green and yellow haploid sectors unlike the parent
strain which only segregates yellow haploids. The genotypes of six green haploids derived from each of these transformants were determined. All were found to be biA1 along with riboB2, while biA1 and riboB+ are on the same chromosome in the untransformed diploid parent. These haploids were also creA+ as determined by growth on sucrose plus allyl alcohol medium. Furthermore, DNA from one of these green haploids was shown by Southern analysis to have the native creA band intact when probed with the insert of pANC7. Therefore, these haploids must have arisen by mitotic recombination events in the diploid transformants rather than being CreA+ transformants. Such an apparently high frequency of mitotic recombination, which is normally a rare event, may be explained by the use of blenoxane as the transformation selection system. The antibiotic bleomycin is cytotoxic because it causes both single and double stranded breaks in DNA (Waring 1981) and it is likely that such an effect may increase the frequency of mitotic recombination events in diploid cells. Bleomycin may therefore be of use in experiments where mitotic recombination is required.

In further attempts to rescue the phenotype of the creA null allele, TC6330 and TC6332 were co-transformed with either pANC7 and pAN7.1 carrying the dominant selectable marker for hygromycin resistance, or pANC7 and p3SR2 containing the amdS gene. Very few Hyg+ and AmdS+ transformants of TC6330 and TC6332 were obtained, the reason for which remains unclear. None of these transformants produced both yellow and green haploid sectors during haploidization analysis and were probably not co-transformants. Further attempts are in progress to obtain larger numbers of transformants of these strains to increase the possibility of identifying stable co-
4.3 SEQUENCES HOMOLOGOUS TO creA FROM A. nidulans:

The creA clone from A. nidulans was used as a probe to DNA samples from various organisms. Southern hybridizations using this probe were performed at low stringency (37°C). Sequences hybridizing to creA were found in DNA from A. niger, A. oryzae, N. crassa, Penicillium chrysogenum and S. cerevisiae. The genome sizes of these fungi are given in the legend to Figure 4.4.1. As shown in Figure 4.4.1, the strongest hybridization was seen to DNA from A. niger with detectable, but lesser amounts of hybridization to the other fungi. It was not unexpected that if creA is an essential wide domain regulatory gene that similar sequences would be found in fungi closely related to Aspergillus since such classes of genes are expected to be highly conserved. This supports further the idea that creA has a function which is central to the mechanism of carbon catabolite repression rather than being specific to a particular metabolic pathway. No hybridization was detected when the creA clone was hybridized at low stringency to DNA from E. coli, Melampsora lini (flax rust), Nicotiana tabacum, Sminthopsis crassicaudata or Homo sapiens. Failure to detect hybridization in these samples may reflect their large genome sizes (listed in the legend to Figure 4.4.1) and/or their evolutionary distance from Aspergillus.

4.4 SUMMARY AND CONCLUSIONS:

The clone which has been isolated from Aspergillus nidulans by marker rescue techniques has been shown to be the region which is
Figure 4.4.1: Southern blot of DNA isolated from various fungi probed with the insert of pANC6 containing the *A. nidulans* creA gene.

Tracks are:

1. 5ug *A. niger* DNA digested with *BamHI*.
2. 15ug *N. crassa* DNA digested with *BamHI*.
3. 5ug *S. cerevisiae* DNA digested with *XbaI*.
4. 5ug *A. niger* DNA digested with *bamHI*.
5. 15ug *A. oryzae* DNA digested with *BamHI*.
6. 15ug *P. chrysogenum* DNA digested with *BamHI*.

Hybridizations were carried out at 37°C using the same conditions outlined in Chapter 2. Autoradiograms were exposed for one week at -80°C using intensifying screens.

No hybridization was detected to *E. coli, M. lini, N. tabacum, S. crassicaudata* or *H. sapiens* DNA samples.

Haploid genome sizes:

*E. coli*       - 4.5 X 10^3 kb (Britten and Kohne 1968).
*S. cerevisiae* - 1.8 X 10^4 kb (Ogur et al. 1952).
*N. crassa*    - 4.3 X 10^4 kb (Horowitz and McLeod 1960).

* S. crassicaudata - 4 X 10^6 kb (Hayman et al. 1982).
*H. sapiens*     - 3 X 10^6 kb (White 1973).
*N. tabacum*     - 2 X 10^4 kb (Bennett et al. 1976).
*M. lini*        - 6 X 10^4 kb (Timmis et al. 1990).

*A. nidulans, A. niger, A. oryzae and P. chrysogenum* probably have similar haploid genome sizes.
disrupted in the creA\textsuperscript{d-30} allele. Analysis of creA\textsuperscript{d-30} shows that this allele is a result of a translocation breakpoint occurring at a position corresponding to the 250bp PstI fragment in the cloned region. The pericentric inversion in creA\textsuperscript{d-30} leads to an altered creA message being expressed, suggesting that creA\textsuperscript{d-30} is not a true loss of function allele. In order to construct a creA gene disruption strain, a larger clone of the region has been isolated. This was used to create a diploid strain containing a creA null allele. As far as can be determined, analysis of this strain demonstrated that a creA null allele is lethal in haploid cells. Although creA appears to be an essential gene and the mutant alleles confer altered creA function, the way in which creA acts at the molecular level remains unclear. Molecular characterization of the creA gene may suggest the role creA has in carbon catabolite repression and whether or not it has features suggestive of a gene encoding a negatively acting repressor or an activator molecule.
CHAPTER FIVE
THE PHYSICAL CHARACTERIZATION OF
THE Aspergillus nidulans creA GENE.

The molecular characterization of the cloned creA gene from *A. nidulans* and computer analyses of the putative CreA protein are described in this chapter. The physical analysis has included the determination of the nucleotide sequence of the creA gene and the structure of its transcript. Computer analyses of the derived amino acid sequence of creA have shown that it contains a number of features characteristic of regulatory and DNA binding proteins. Amino acid sequence comparisons between the putative CreA protein and other proteins of known function lead to suggestions as to how CreA functions at the molecular level.

5.1 THE PHYSICAL ANALYSIS OF creA:

5.1.1 Nucleotide sequence analysis of creA and characterization of the coding region:

Subclones of a 2.7kb region starting at the internal BamHI site in pANC6 and spanning the creA gene were sequenced using the strategy outlined in Figure 5.1.1. The direction of transcription of the creA gene is from the internal BamHI site in pANC6 towards the XbaI site (as reported in Chapter 3, section 3.7). The complete nucleotide sequence of this 2.7kb region containing the creA gene is shown in
FIGURE 5.1.1: Partial restriction map of the insert in pANC6 showing the strategy used for the sequencing of the *A. nidulans creA* gene and its flanking regions. The extent and direction of sequence obtained from the insert of pANC6 is indicated by the arrows below the restriction map. Sequence data obtained from cDNA clones is also shown. The positions of the start (ATG) and end (TAA) of translation are indicated by the vertical arrows.

Restriction site abbreviations: 

- **B** - *BamHI*  
- **E** - *EcoRI*  
- **Ev** - *EcoRV*  
- **H** - *HindIII*  
- **Ps** - *PstI*  
- **Pv** - *PvuII*  
- **S** - *SalI*  
- **Sc** - *Sacl*  
- **X** - *XbaI*
In order to identify the coding region of the creA gene, twelve cDNA clones were isolated from the cDNA libraries described in Chapter Two. These cDNA clones were isolated either using the entire BamHI-XbaI insert of pANC4 or the 1.2kb BamHI-PstI fragment containing the 5' region of the gene as probes to plaque lifts of these libraries. In total, seven cDNA clones of different lengths were isolated. Two of these, designated pCD1 and pCD5 were subcloned and sequenced and found to span the entire genomic sequence from nucleotide position -599 through to 1602. Therefore, sequence analysis of these cDNA clones demonstrates that the creA gene does not contain any introns. The region covered by these cDNA clones is indicated in Figure 5.1.2. Furthermore, a computer search failed to reveal any putative intron splice sites or consensus lariat sequences which would allow intron excision and the maintenance of the reading frame. Although the majority of genes analysed from A. nidulans and other filamentous fungi have been shown to contain short introns of less than 100bp in length, a number contain no introns (reviewed by Gurr et al. 1987) including the gene involved in ammonium repression, areA, which has a single long ORF in the sequenced region of 2157bp in length (Kudla et al. 1990).

The entire sequence was compared to nucleic acid data bases to determine if any sequence similarities could be found between the creA sequence and other genes of known function. No significant sequence similarities could be identified although overall sequence similarity was shown between the creA sequence and that of the brlA gene from A. nidulans. However, more precise alignments of these two sequences failed to show any extended stretches of homology and the
FIGURE 5.1.2: Nucleotide sequence of the *A. nidulans* creA gene and its flanking regions. Numbering of nucleotides begins with the first nucleotide of the start codon as +1. Putative TATAA and CAAT boxes are underlined. The major start point of transcription is indicated by a solid wide arrow. The extent of cDNA clones pCD1 and pCD5 are shown by small arrow heads and wide open arrows respectively. The 3' end of pDC5 also denotes the position of polyadenylation of the creA transcript. Possible signals for polyadenylation are indicated with asterisks. The putative amino acid sequence is shown below the corresponding codons. The boxed regions contain the two putative zinc fingers and an alanine rich region is indicated by stars under each of the residues. The sequence is continued overleaf.
Phe Ser Asn Tyr Ala Asn His Met Arg Ser Asn Leu Ser Pro Tyr Ser Arg

592 ACC AGT GAA CGG GCG TCA TCA GGC ATG GAT ATC AAC CTT CTT GCT ACG GCC
Thr Ser Glu Arg Ala Ser Ser Gly Met Asp Ile Asn Ala Thr Ala

643 GCG TCT CAA GTG GAT GAA AGT TTT GGA TTC CGC TCT GGT CAA CGT
Ala Ser Gin Val Glu Arg Asp Glu Ser Phe Gly Phe Arg Ser Gly Gin Arg

694 AGT CAC CAT ATG TAT GGT CCC CGC CAT GGC AGC AGG GGA CTT CTT
Ser His His Met Tyr Gly Pro Arg His Gly Ser Arg Gly Pro Ser Leu

745 TCA GCC TAC GCC ATC TCC CAC AGC ATG AGC CTT CAT TCC CAT GAC GAG GAT
Ser Ala Tyr Ala Ile Ser His Ser Met Ser Arg Ser His Ser His Glu Asp

796 GAG GAT TCT TAT GCG TCA CAT CGC GTC AAG CGT TCA AGA CCT AAC TCA CCC
Glu Asp Ser Tyr Ala Ser His Arg Val Lys Arg Ser Arg Pro Ser Pro

847 AAC TCG ACT GCT CCT TCT TCG CCT ACC TTC CAC GAC TCC TTA TCT CCC
Asn Ser Thr Ala Pro Ser Ser Thr Pro Thr Ser His Asp Ser Leu Ser Pro

898 ACT CCT GAC CAC ACG CCA TTG GCT ACG CCC GCC CAT TCG CCA CGA CTG AAG
Thr Pro Asp His Thr Pro Leu Ala Thr Pro Ala His Ser Arg Leu Lys

949 CCA TTG TCG CCG AGT GAG CTA CAT CTG CCC TCA ATC GCT CTA TCG CTT
Pro Leu Ser Pro Ser Glu Leu His Leu Pro Ser Ile Arg His Leu Ser Leu

1000 CAC CAC ACT CGG CTG CTC GTT CCA ATG GAG CCC CAG GCC GAG GGA CCC AAT
His His Thr Pro Arg Leu Arg Met Glu Pro Gin Ala Glu Gly Pro Asn

1051 TAT TAT AAC CCG AAC CAA CCT CAT GGT GGC CCA AGC ATA AGC GAT ATC
Tyr Tyr Asn Pro Asn Gin Pro His Val Gly Pro Ser Ile Asp Arg Met

1102 TCT CGC CCT GAG CTG CAC AGC GAA AAC TTC CGA TAC CTC AGG TGC CCA AAG
Ser Arg Val His Ser Glu Asn Phe Arg Tyr Leu Arg Cys Pro Lys

1153 TGG CGG TCC AGG ATA TGT TAA ATCCTAGCGGGTTTACTTCTCATTCAACCGCAA
Trp Arg Ser Arg Ile Cys End

1213 ATTCGTTGCTGGTGGTACTGGTGAAGGTTTCTAATCGGGCAAAAAACCTTCGGTTTTTCTGTGA

1280 GGCCTAGAAGATATAGACCTTTTGCAATTTCTGGTTGATTGCTAGCATTTGGCTGACGGAAT

1347 AGGTGTTTGGACGATTCTTCACACTTGGTGTGATACATATATTTTGGCGAGGCGGTGTCCTCCTATAGACGG

1414 ATTATCCCTGGCCTTCACACAAAGTCCTTGTTTTTATCTACTTTCACTCATTCTTTCTCTTCTCTCTCAGCTACTTT

1481 ACACCAACGTATATCTCTTCTCTGAGATCTCCTCTCTCTGACAAAGCGATTCGCCTCAGCAGTTTACT

1548 ATTTCTTGGGTTACTACCTCATATAACTCAAAATACAAAGGAGGTGGTTTCTGTGAAATGCAACACTGTG

1615 CCAATTTACCTTCTTCTCGATTTAAATGAATACTCCCTCATGAGAGCACTACAGATGCATTCCCCCTCAATAGACGG

1682 GCGCACTATGAGACCCCCGCAATGTCCTCCAGAGCACTCCCATCCTCATATCGCATGTTATACAGCGCCT

1749 GCCGTAGAATCGTCTGAGTCTCATTGTTGACTCTTATCCTTACTCTTACTGTAAGAGGCTTATGTTATATCTCCAT

1816 CTGGAAGGTGTAGATACTGAGTCTATAGACACTTTGCTCAATATC
similarity probably results from the high frequency of serine residues in the two encoded proteins.

Codon usage within the creA gene is shown in Table 5.1.1. Many genes from fungi, especially those that are highly expressed, have been shown to have a bias against A and for C in the third position of codons (reviewed by Ballance 1986, reviewed by Gurr et al. 1987). The codon usage in creA is not random and some degree of codon bias is present. However, codon bias is not marked in the creA coding region which may be a reflection of the low levels of expression of this gene in A. nidulans. Low codon bias is also found in the coding regions of the regulatory genes areA from A. nidulans (Kudla et al. 1990) and nmr from N. crassa (Young et al. 1990) and therefore may be more a feature of regulatory rather than of structural genes from filamentous fungi. Furthermore, reduced codon bias has also been correlated with those fungal genes for which there is a low turnover of both its transcript and the encoded protein within cells (reviewed by Gurr et al. 1987). AGN codons for serine and arginine have been found to be used infrequently in genes from filamentous fungi. In the creA gene there are 22 AGN codons out of 391. This frequency is unusually high with the CreA protein having a high serine content in which 16 out of the 59 serine residues are encoded by AGN codons. Filamentous fungal genes show a preference for the termination codon TAA and this is used for the termination of translation in the creA transcript.

5.1.2 The 5' region of the creA gene:

The major start point of transcription was mapped by primer
Table 5.1.1: Codon usage in the creA gene of *A. nidulans*.

<table>
<thead>
<tr>
<th>Gly</th>
<th>GGG 1</th>
<th>Ser</th>
<th>AGT 7</th>
<th>Cys</th>
<th>TGT 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGA</td>
<td>7</td>
<td></td>
<td>AGC 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGT</td>
<td>4</td>
<td></td>
<td>TCG 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCC</td>
<td>5</td>
<td></td>
<td>TCA 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>GAG 10</td>
<td></td>
<td>TCC 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAA</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>GAT 8</td>
<td>Lys</td>
<td>AAG 8</td>
<td></td>
<td>Leu 3</td>
</tr>
<tr>
<td></td>
<td>GAC 4</td>
<td></td>
<td>AAA 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CTG 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>GTG 3</td>
<td>Asn</td>
<td>AAT 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GTA 0</td>
<td></td>
<td>AAC 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GTC 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>GCG 7</td>
<td>Met</td>
<td>ATG 12</td>
<td>Phe</td>
<td>TTT 1</td>
</tr>
<tr>
<td></td>
<td>GCA 4</td>
<td></td>
<td>ATT 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCT 15</td>
<td></td>
<td>ATC 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCC 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>AGG 3</td>
<td>Ile</td>
<td>ATA 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGA 3</td>
<td></td>
<td>ATT 2</td>
<td>Gln</td>
<td>CAG 3</td>
</tr>
<tr>
<td></td>
<td>CGG 3</td>
<td></td>
<td>ATC 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGA 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGT 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGC 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Thr 4</td>
<td>His</td>
<td>CAT 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>AGG 3</td>
<td>Pro</td>
<td>ACT 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGA 3</td>
<td></td>
<td>ACC 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGG 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGA 5</td>
<td></td>
<td>Trp 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGT 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGC 10</td>
<td></td>
<td>End 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TAG 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TAA 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


extension analysis (Figure 5.1.3) and was positioned at nucleotide -589. Although some fungal genes have been shown to have a unique start site for transcription initiation, many genes have multiple initiation sites. Evidence of 5' heterogeneity for the creA transcript is shown by the difference between the major start point as mapped by primer extension analysis and the 5' end of the cDNA clone pCD5 which begins at nucleotide position -604 in the genomic sequence. However, although pCD5 is probably a full length cDNA clone it has an anomalous tract of more than thirty thymidine residues upstream of the 5' end of homology to the genomic sequence. This can not be accounted for by the presence of an intron which brings together the long stretches of oligo (d)T present between nucleotides -788 and -658 in the genomic sequence. Furthermore, it is not possible that this represents a polyA tail at the 3' end of a gene since long ORF's are not present in the complement of the sequence of pCD5. It was therefore assumed that this region was an artifact created during the construction of this cDNA library. Presumably, the start site determined from primer extension analysis represents the major start point of transcription and other minor start sites, including the one represented by the 5' end of pCD5, may have been indicated by longer exposures of the primer extension products. The region between the putative TATAA-like sequence and the major start point of transcription is pyrimidine rich where 20/20 bases are pyrimidines. This is also seen in the promoters of other fungal genes, particularly in genes from S. cerevisiae which are highly expressed. Pyrimidine rich promoter sequences are also found to be common in highly expressed genes from Aspergillus (Ward and Turner 1986, Clements and Roberts 1986, Punt et al. 1988, Ward et al. 1988). The major start point of transcription of the A. nidulans creA gene
FIGURE 5.1.3: Primer extension analysis for the mapping the 5' end of the *A. nidulans creA* mRNA transcript. A γ-32P-labelled oligonucleotide primer of the sequence TCTAAACAAGGGTCTAAA, complementary to nucleotides -454 to -436 (Figure 5.1.2) was annealed at 43°C to total mRNA isolated from wildtype *A. nidulans* mycelia from glucose grown cultures and extended with M-MLV reverse transcriptase by the method of Geliebter *et al.* (1986) as described in Section 2.7.10. The products of the extension (tracks 1-3) and of a dideoxynucleotide sequencing reaction (tracks T, C, G and A) were electrophoresed on a 6% polyacrylamide sequencing gel (Section 2.7.5). The clone sequenced for the molecular weight marker was pANC7 primed with the same oligonucleotide as above.

The tracks are: 1. Primer extension reaction using 20μg total mRNA.
    2. Primer extension reaction using 10μg total mRNA.
    3. Primer extension reaction without mRNA.

The length of the extended product was 154 nucleotides which allows the positioning of the major start point of transcription at -589 in the nucleotide sequence (Figure 5.1.2).

Exposure of the extension reactions to X-ray film was carried out at -80°C using an intensifying screen.
is located at the first purine after this pyrimidine rich region which is the case for other Aspergillus genes. Deletion of a small pyrimidine rich region of the A. nidulans trpC promoter results in transcription initiating from heterogeneous positions (Hamer and Timberlake 1987). Therefore, this sequence may play a role in directing transcriptional start points rather than affecting high levels of expression.

The 5' untranslated region of the creA transcript is 589bp in length which is relatively long for fungal genes. Although considerable variation exists in the length of the 5' untranslated mRNA of fungal genes, most commonly this region is 30-70bp in length (reviewed by Gurr et al. 1987). However, other genes from A. nidulans have been shown to have unusually long 5' untranslated regions, for example, analysis of a cDNA clone for the gene bimG from A. nidulans demonstrated that its mRNA has an untranslated 5' region that is 885 bases long (Doonan and Morris 1989). As more genes from fungi are characterized it is becoming evident that greater variation exists in the length of mRNA untranslated regions.

A TATAA-like sequence is present at about 30bp upstream from the major start point of transcription in many fungal genes (reviewed by Gurr et al. 1987). The promoter region of the creA gene contains a TATAA-like sequence beginning at -26 to the major start point of transcription. Furthermore, another A-T rich sequence is present at -731 in the genomic sequence (142bp upstream from the start point of transcription) resembles a TATAA box. The CAAT motif is also present in the promoters of many eukaryotic genes (reviewed by Gurr et al. 1987). These are usually at about -70 to -90 upstream from the start point of transcription. Multiple CATT box sequences are present at
62bp upstream from the major start point of transcription in the creA sequence. However, the role of TATAA and CAAT boxes in the promoter regions of fungal genes remains unclear and whether or not they affect expression levels of the genes in which they occur can only be determined from mutational analyses of these promoters. Deletions of TATAA boxes from yeast genes have demonstrated that the resulting constructs have severely reduced levels of transcription without affecting regulation. This suggests that the TATAA box may represent at least part of a promoter element in yeast genes (reviewed by Guarente 1984). This is probably not entirely the case for all promoters from filamentous fungi. Such sequences have not been found in many genes from these fungi and may be deleted from some filamentous fungal promoters without affecting expression (Hamer and Timberlake 1987).

The preferred bases immediately preceding the start codon in most fungal genes are TCA in positions -5 to -3 (Ballance 1986). The sequence immediately upstream from the start codon in creA (TTCACAATG) resembles that of other fungal genes although the TCA is located at nucleotide positions -4 to -2.

The 5' region of the creA gene contains large blocks of C-T rich sequence. This includes long stretches of T's, the longest being a run of 22 T's. The significance of these 5' sequences is unclear although C-T rich regions are found in many fungal promoters. In eukaryotes, the upstream regions of genes often have altered chromatin structures (Nussinov et al. 1984, Nussinov 1985, reviewed by Nussinov 1990). Physical chemical studies have suggested that oligo (dA-dT) tracts in DNA sequence can form unique conformations and cause DNA bending. Extreme difficulty was experienced in
sequencing through the oligo d(T) stretches in the creA upstream region, characterized by large compressions and pauses on sequencing gels. This may have been due to secondary structures in the DNA caused by the oligo d(T) runs. Similar regions containing oligo (dA-dT) tracts have been located upstream from promoter elements (Bossi and Smith 1984, Plaskon and Wartell 1987) and within promoters in protein binding sites and origins of replication (Koepsel and Khann 1986, Snyder et al. 1986, Zahn and Blattner 1987, Inokuchi et al. 1988). Studies on the interaction between T-antigen binding and SV40 have demonstrated that sequences flanking oligo (dA-dT) tracts are bound to this protein. The oligo (dA-dT) tract is itself not involved in protein binding but instead is required for bringing the flanking regions into the proper position for T-antigen binding (Maroun and Olson 1988). Oligo (dA-dT) tracts have also been found to occur frequently in intergenic regions in the genome of S. cerevisiae and have been shown to function, in either orientation, as promoter elements (Struhl 1985a, 1985b). These elements may function by association with the nuclear scaffolding in yeast cells (Gasser and Laemmli 1986) or by interfering with nucleosome formation (Kunkel and Martinson 1981, Prunell 1982). Other reports suggest that oligo d(T) stretches on the DNA coding strand may act as UAS's by excluding DNA binding proteins (Russell et al. 1983, Chen et al. 1987) or by acting as protein binding sites themselves (Lue et al. 1989, Thiry-Blaise and Loppes 1990). Recently it has also been demonstrated that a single oligo (dA-dT) sequence is the site for the initiation of transcription of two divergently transcribed yeast genes (Schlapp and Roedel 1990).
5.1.3 The 3' region of the creA gene:

Like the 5' untranslated region of the creA transcript, the 3' untranslated region is also very long, extending 428bp from the 3' end of cDNA clone pCDS. Although extensive 3' untranslated sequences are uncommon in genes from A. nidulans, they are not unknown. For example, the 3' untranslated portion of the areA transcript is reported to be 539 bases long (Kudla et al. 1990). The site of polyadenylation of the creA transcript was determined by the presence of a small poly(A) tail of six adenosine residues at the 3' end of pCDS. The addition of poly(A) usually occurs approximately 20bp downstream of the consensus sequence AATAAA or related sequences such as ATAA or AATA in fungal genes (reviewed by Gurr et al. 1987). These smaller core sequences occur in the 3' region of the creA gene 25 and 32bp upstream from the site of polyadenylation and may be the signals for polyadenylation of the creA transcript.

5.2 ANALYSIS OF THE PUTATIVE CreA PROTEIN:

The predicted CreA protein is 390 amino acids in length (Figure 5.1.1) and exhibits several features characteristic of regulatory and DNA binding proteins. These structural motifs are discussed in the following sections along with the general features of the protein as predicted from computer analyses. Comparison of the entire CreA protein sequence with the GenEMBL database showed that the top one hundred best score protein sequences identified as having sequence similarity to CreA were almost exclusively sequences containing a zinc finger DNA binding motif. The results of further more detailed sequence comparisons are discussed below.
5.2.1 The zinc finger motif:

The derived CreA protein sequence contains two putative zinc finger DNA binding motifs (Klug and Rhodes 1987) that closely resemble the C₂H₂ zinc fingers first recognized in the transcription factor TFIIIA from Xenopus laevis (Miller et al. 1985). A large number of proteins from a variety of eukaryotic organisms have subsequently been found to contain regions that resemble the zinc finger of TFIIIA. Some of these proteins are known to be transcriptional activators or function as repressors (reviewed by Levine and Manley 1989, Nehlin and Ronne 1990). The first zinc finger in CreA begins at residue 64 in the amino acid sequence and is closely followed by a second zinc finger beginning at residue 92 (Figure 5.1.1). Figure 5.2.1 demonstrates the striking amino acid similarity between the two zinc fingers in CreA and other zinc finger domains from a variety of proteins and suggests that CreA is a DNA binding protein. The consensus sequence for C₂H₂ zinc fingers is Cys-X₂₋₄-Cys-X₃-Phe-X₅-Leu-X₂-His-X₃₋₄-His, along with conserved flanking residues and is characterized by specifically spaced, conserved amino acid residues that are arranged so that the cysteine and histidine residues are tetrahedrally co-ordinated to a single atom of zinc. Both these zinc fingers are of the C₂H₂ type and their predicted structures based on the model for the conformation of these sequences are shown in Figure 5.2.2. There is evidence suggesting that this organization allows the conserved hydrophobic phenylalanine and leucine residues to form hydrophobic interactions thereby giving stability to the tip of the finger (reviewed by Bray and Thiesen 1990). The first zinc finger in CreA does not have this usually conserved leucine residue but instead has a glutamine at this
FIGURE 5.2.1: Alignment of the putative CreA zinc finger domains with the amino acid sequences of other proteins containing the \( \text{C}_2\text{H}_2 \) zinc finger motif. Numbers after the proteins specify the number of the finger represented. Arrows indicate the cysteine and histidine residues of the \( \text{C}_2\text{H}_2 \) motifs.

The protein sequences are:

- **TFIIIA** - Transcription factor IIIA from *X. laevis* (Miller *et al.* 1985).
- **BrlA** - Protein involved in developmental regulation from *A. nidulans* (Adams *et al.* 1988).
- **WT** - Protein encoded by a gene which is deleted in Wilms' tumor (Call *et al.* 1990).
- **ADRI** - Regulator of *ADH2* from *S. cerevisiae* (Hartshorne *et al.* 1986).
- **Kr** - Kruppel from *D. melanogaster* (Rosenberg *et al.* 1986).
- **MIG1** - Repressor of *SUC2* gene from *S. cerevisiae* (Nehlin and Ronne 1990).
FIGURE 5.2.2: Amino acid sequence of zinc fingers I and II of CreA drawn to illustrate the finger motifs as proposed by Miller et al. (1985). Highly conserved residues are circled.
position. Whether or not the structure of the first zinc finger in CreA is affected by this difference is not clear. However, the first zinc finger in the MIG 1 protein from yeast also has a glutamine in this position. Furthermore, the alignment of the amino acid sequences of CreA and MIG 1 show striking similarities between the entire zinc finger regions of both proteins (Figure 5.2.3). MIG 1 is required for glucose repression of the suc2 gene from S. cerevisiae and also like the creA gene, encodes a protein containing two consecutive zinc fingers. The zinc fingers in this repressor protein were identified as being very similar to the zinc fingers of the human early growth response proteins (Egr 1 and 2) and the zinc finger domains in a protein encoded by a gene which is deleted in Wilms tumor. MIG 1 has been shown to bind to two target sites approximately 500bp upstream from the suc2 gene (Nehlin and Ronne 1990). That CreA binds to specific DNA sequences via its putative zinc fingers and whether both domains are involved in binding remains to be confirmed.

It seems clear that TFIIIA-like zinc fingers are responsible for sequence specific DNA binding and that this activity requires the presence of more than one finger domains in tandem (or perhaps dimers) of single finger proteins. For example, Br1A which is involved in the transcriptional regulation of developmental genes in A. nidulans, has two zinc fingers which are both required for activity of the protein (Adams et al. 1990). Another conserved region in zinc finger proteins is the sequence between two successive zinc finger domains called the H-C link which spaces repeated finger domains along the DNA helix. The H-C link in the CreA protein shows some sequence similarity to the consensus sequence TGEKPHA (reviewed by Bray and Thiesen 1990) (Figure 5.2.4).
FIGURE 5.2.3: Comparison of the amino acid sequences of the zinc finger regions from CreA (this thesis) and MIG1 (Nehlin and Ronne 1990). Double dots indicate the same amino acids are present in each sequence while single dots show similar amino acids. Arrows indicate the histidine and cystine residues of the C2H2 motifs.

FIGURE 5.2.4: A comparison of the H-C link regions of zinc fingers from various proteins with the same region from CreA. Lower case letters indicate differences from the consensus (Bray and Thiesen 1990). The Kruppel sequence is the consensus found for this protein (Rosenberg et al. 1986) where X denotes any amino acid.

The protein sequences are:

**BrlA** - Protein involved in developmental regulation from *A. nidulans* (Adams et al. 1988).

**Kruppel** - Kruppel from *D. melanogaster* (Rosenberg et al. 1986).

**Mouse MKR** - Mouse growth factor induced protein (Christy et al. 1988).

**MIG1** - Repressor of SUC2 expression from *S. cerevisiae* (Nehlin and Ronne 1990).

**Sp1** - Human transcription factor Sp1 (Kadonaga et al. 1987).

**HPFI** - Human zinc finger protein HPFI (Bellefroid et al. 1989).

**ZFP7** - Human zinc finger protein 7 (Lania et al. 1990).


**MFG1** - Mouse zinc finger protein (Passananti et al. 1989).
CREA
PRPYCPLCERAFAQHRLEHQTRHIRTHTGEKPHACQPFGCSKRFERSDELTRHRSRIHNN

MIG1
PRPHAPICHRFAQHRLEQTRHRMRIHTGEKPHACDFPGCVKRFERSDELTRHRRIHNN

CREA H T G E K P h A C
BRLA H s k E K P h V C
KRUPPEL H T G E K P Y X C
MOUSE MKR H T G E K P Y E C
MIG1 H T G E K P h A C
Sp1 1 H T G E r P f M C
2 H T G E K k f A C
HPF1 H T G E K P Y K C
ZFP7 H T G E r P Y P C
ZFP-36 H T G E K P Y E C
MFG1 H T G E K P Y K C

CONSENSUS H T G E K P Y X C
There have been no reports of the successful crystallization of an intact zinc finger, although single finger domains have been analysed. Predictions of molecular structure have arisen from two dimensional and three dimensional NMR studies of finger fragments and from computer modelling of DNA-protein interactions (Berg 1988, Lee et al. 1989). These models indicate that most fingers form a β-turn - α-helical structure where non-specific interactions occur between positively charged amino acid side chains and phosphates of the DNA backbone. Sequence specific interactions result from the recognition of two nucleotide pairs by five amino acid residues situated on the exposed helical surface and NH₂ terminus of the finger. The predicted peptide structural features of the region spanning the two CreA zinc fingers is shown in Figure 5.2.5. The predictions show that these sequences may also have structural similarities as well as differences to characterized zinc finger domains. Some fingers such as fingers 1, 3 and 6 of TFIIB have significant structural differences as predicted by protein modelling (reported by Bray and Thiesen 1990). However the functional constraints of these different structures in these fingers and the corresponding domains in CreA are unclear. The probability of the amino acids spanning the zinc fingers in CreA being on the surface of the folded protein is shown in Figure 5.2.5. The tips of the fingers (shown in Figure 5.2.2) occur between residues 70-81 in zinc finger I and 100-111 in zinc finger II. The amino acids in these regions are predicted to be on the surface of the protein due to their hydrophobic nature and this is consistent with the likelihood that these regions do form the tips of functional zinc fingers. The tips of zinc fingers are important for DNA binding and the core amino acid residues in these regions show variation between different zinc finger proteins and are
FIGURE 5.2.5: Plot structure of Peptidestructure (Devereux 1984) for CreA amino acid sequence spanning the two zinc fingers from residue 50 to 150. Residues are numbered from 1 to 100 on the X axis.

Predictions are:

**KD** Hydrophilicity - hydrophilicity according to Kyte and Doolittle (1982).

**Surface Prob.** - surface probability according to Emini *et al.* (1985).

**Flexibility** - flexibility of peptide chain according to Karplus and Schulz (Devereux 1984).

**Jamson-Wolf Antigenic index** - a measure of the probability that the region of the protein is antigenic (Devereux 1984).

**CF Turns** - residue conformations most commonly found in turns according to Chou and Fasman (1978).

**CF Alpha Helices** - Chou and Fasman prediction of alpha-helix forming regions that are not in conflict with other secondary structures (Chou and Fasman 1978, Nishikawa 1983).

**CF Beta Sheets** - Chou and Fasman prediction of beta-sheet structures that are not in conflict with other secondary structures (Chou and Fasman 1978, Nishikawa 1983).

**GOR Turns** - predictions of turns in protein according to Garier-Osguthorpe-Robson (Garnier *et al.* 1978).

**GOR Alpha Helices** - Garier-Osguthorpe-Robson prediction of alpha-helix forming regions (Garnier *et al.* 1978).

**GOR Beta Sheets** - Garier-Osguthorpe-Robson prediction of beta sheet structures (Garnier *et al.* 1978).

**Glycosyl. Sites** - predicted sites for glycosylation where the residues have the composition NXT or NXS (Cohen *et al.* 1984).
presumed to give zinc finger proteins their specificity to target sites (Miller et al. 1985, Bellefroid et al. 1989). Comparison of the amino acid residues within the tips of the zinc fingers from MIG 1 with other zinc fingers indicated that the central residues FSRSD in zinc finger II are identical in a number of other zinc fingers (Nehlin and Ronne 1990). An alignment of these zinc finger loops to the same region in zinc finger II of CreA is shown in Figure 5.2.6. All (except Sp1) contain the central residues FSRSD and a high degree of conservation of flanking amino acids is also evident. The sequences to which some of these proteins bind (Figure 5.2.6) have been identified and their conservation has been noted (Chavrier 1990, Nehlin and Ronne 1990, Nardelli et al. 1991). It will therefore be of interest to determine whether CreA also binds to a similarly related target sequence.

5.2.2 An alanine rich region:

The putative CreA protein sequence was compared to protein sequences known to have transcriptional activation or repression functions as well as to many other regulatory proteins from various organisms. This was done in an attempt to define other domains within the CreA protein which may indicate how CreA functions at the molecular level.

Along with regions of amino acid similarity due to the presence of zinc fingers, the alanine rich region occurring after the second zinc finger in CreA was identified as being similar to domains present in a number of regulatory proteins. The alanine rich region in CreA consists of nine consecutive alanine residues beginning at
FIGURE 5.2.6: Comparison of zinc finger loop regions from CreA with other similar zinc finger loops from various zinc finger proteins. Numbers refer to which zinc finger from the protein is given. Target sequences to which some of these proteins bind are also shown (references appear in the text).

Sequences are from the following proteins:

- **BrlA** - Protein involved in developmental regulation from *A. nidulans* (Adams *et al.* 1988).


- **Mouse EGR1** - Mouse early growth response protein 1 (Sukhatme *et al.* 1988).

- **Mouse GFI** - Mouse growth factor induced protein (Christy *et al.* 1988).


- **Krox-20** - Mouse serum inducible transcriptional activator (Chavrier *et al.* 1988).


- **MIG1** - Repressor of *SUC2* expression from *S. cerevisiae* (Nehlin and Ronne 1990).

- **Spl** - Human transcription factor Spl (Kadonaga *et al.* 1987).
<table>
<thead>
<tr>
<th>PROTEIN</th>
<th>ZINC FINGER LOOP</th>
<th>BINDING SEQUENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>CREA 1</td>
<td>CERAFHRLEHQTTRH</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Cskrfssrdeltrh</td>
<td></td>
</tr>
<tr>
<td>BRLA 2</td>
<td>Chrafssrdelah</td>
<td></td>
</tr>
<tr>
<td>HUMAN EGR1</td>
<td>CDRFSSRSDELTRH</td>
<td></td>
</tr>
<tr>
<td>HUMAN EGR2</td>
<td>CDRFSSRSDELTRH</td>
<td></td>
</tr>
<tr>
<td>MOUSE EGR1</td>
<td>CDRFSSRSDELTRH</td>
<td></td>
</tr>
<tr>
<td>MOUSE GFI</td>
<td>CDRFSSRSDELTRH</td>
<td></td>
</tr>
<tr>
<td>NERVE GF</td>
<td>CDRFSSRSDELTRH</td>
<td></td>
</tr>
<tr>
<td>Krox-20 1</td>
<td>CDRFSSRSDELTRH</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krox-24</td>
<td>CDRFSSRSDELTRH</td>
<td></td>
</tr>
<tr>
<td>MIG1 1</td>
<td>Chrafhrlehqttrh</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CVKRFSRSDELTRH</td>
<td></td>
</tr>
<tr>
<td>Sp1</td>
<td>CGKRFTSRDELQRHR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
position 131 in the amino acid sequence while a significant number of alanine residues are also present upstream of the zinc finger region (Figure 5.2.1). Similar alanine rich regions occur in the repressor proteins kruppel, engrailed and evenskipped from Drosophila melanogaster (Figures 5.2.7-5.2.9). kruppel acts as a negative regulator of transcription in D. melanogaster and represses the expression of the pair-rule gene evenskipped (Goto et al. 1989) and the gap gene hunchback (Jackle 1986) during the blastoderm stage of D. melanogaster development. It has been demonstrated that the DNA binding function and transcriptional repression activities are separate domains in the kruppel protein, as they are in most DNA binding transcription factors. The repression function of kruppel has been mapped to the NH₂-terminal region of the protein showing the similarity to the alanine rich region in CreA. Comparisons of the repressor domain of kruppel to the amino acid sequences of evenskipped and engrailed identified similar alanine rich regions in these proteins. It has also been reported that the deletion of the alanine rich region of evenskipped leads to the protein having a reduced ability to act as a repressor of transcription (Han and Manley reported by Licht et al. 1990). It is therefore not surprising that these proteins also showed homology to the alanine rich region in CreA. Significant similarity between CreA and engrailed (Figure 5.2.8) extends past the alanine rich region and probably reflects the high serine content of these regions. Alanine residues are relatively hydrophobic and therefore probably promote the formation of α-helices. Alanine rich regions may therefore be important for protein-protein interactions.

A number of eukaryotic transcription factors, including kruppel,
FIGURE 5.2.7: Compare program output between putative CreA protein and Kruppel protein from *D. melanogaster*. Points are marked where the two amino acid sequences are similar within the window size indicated (Maizel and Lenk 1981, Devereux 1984). The circled region shows the position of an alanine rich region in both sequences.
Kruppel.Pep ck: 5,834, 1 to 468
FIGURE 5.2.8: Compare program output between putative CreA protein and Engrailed protein from D. melanogaster. Points are marked where the two amino acid sequences are similar within the window size indicated (Maizel and Lenk 1981, Devereux 1984). The circled region shows the position of an alanine rich region in both sequences.
FIGURE 5.2.9: Compare program output between putative CreA protein and Evenskipped protein from D. melanogaster. Points are marked where the two amino acid sequences are similar within the window size indicated (Maizel and Lenk 1981, Devereux 1984). The circled region shows the position of an alanine rich region in both sequences.
have been shown to have both activator and repressor functions. Genetical data indicates that kruppel acts to positively regulate another gap gene, knirps, and further to this, the knirps promoter has been found to contain a binding site for the kruppel protein (Pankratz et al. 1989). However, the activator domain of kruppel is yet to be defined and mapped (Licht et al. 1990). That activator and repressor functions of the ultrabithorax protein from D. melanogaster can be separated (Krasnow et al. 1989) suggests that this may be the case for other eukaryotic transcription factors having both activities (reviewed by Levine and Manley 1989). Therefore, although these protein domain similarities suggest that CreA could have repressor function at the level of transcription, it can not be ruled out that it may also have positively acting regulatory functions involving the activation of transcription of some other (as yet genetically unidentified) genes controlled by CreA.

Cyc8 (SSN6) acts as a negative regulator of gene expression during glucose repression in S. cerevisiae (Trumbly 1986) and like the above mentioned D. melanogaster genes it also has an alanine rich region (Trumbly 1988) showing homology to the alanine rich region of CreA (Figure 5.2.10). However, deletion studies of the Cyc8 protein have indicated that this region is not important for at least some of the activities of Cyc8, although the functional significance of other glutamine/alanine rich regions in this protein are yet to be analysed (Schultz et al. 1990). The importance and/or significance of this region for CreA function therefore remains to be determined through the functional analysis of truncated CreA proteins.
FIGURE 5.2.10: Compare program output between putative CreA protein and Cyc8 protein from *S. cerevisiae*. Points are marked where the two amino acid sequences are similar within the window size indicated (Maizel and Lenk 1981, Devereux 1984). The circled region shows the position of an alanine rich region in both sequences.
5.2.3 The S/TPXX motifs:

The amino acid sequence motifs Ser-Pro-X-X and Thr-Pro-X-X (S/TPXX) are found to occur more frequently in regulatory proteins such as the *Drosophila* homeotic proteins and steroid hormone receptors than in other DNA binding proteins such as the core histones which are not directly involved in gene regulation. It has been suggested that these motifs assume a β-turn structure and contribute to the DNA binding capacity of proteins (Suzuki 1989). The putative CreA protein contains 58 serine and 43 proline residues and eight SPXX and four TPXX motifs. The expected frequency (calculated as in Suzuki 1989) for SPXX and TPXX motifs occurring at random in the CreA protein are $1.6 \times 10^{-2}$ and $4.5 \times 10^{-3}$ respectively, while the observed frequencies are $2.0 \times 10^{-2}$ and $1.0 \times 10^{-2}$ respectively. Thus, these motifs occur more frequently in the CreA protein than expected and more frequently than is found in general proteins ($2.89 \times 10^{-3}$ and $3.08 \times 10^{-3}$ respectively) and may reflect the regulatory nature of the protein. AreA is another zinc finger protein from *A. nidulans* and like CreA, it also has a high frequency of these motifs (Kudla *et al.* 1990).

5.2.4 PEST sequences:

It has been reported that many proteins with intracellular half lives of less than two hours contain regions (PEST sequences) which are rich in proline(P), glutamic acid(E), serine(S) and threonine(T) and lack internal hydrophobic residues which are flanked by several positively charged amino acids. Since the expression of the creA gene appears to be constitutive, it is possible that other mechanisms such
as those affecting protein stability assist in regulating the activities of CreA within the cell. A computer program generated by Rogers et al. (1986) was used to scan the CreA amino acid sequence for PEST sequences. Four PEST-like regions were identified using this program of which two, located between positions 1-47 and 163-179, had PEST-SCORES of -5.5 and 0.97 respectively, which are similar scores to those reported for proteins having short intracellular half lives. However, the relative importance of these sequences, with respect to other structural features of proteins, in affecting stability remains to be clearly understood. No experimental evidence has been reported to suggest the involvement of PEST sequences in protein stability, however they may be useful for predicting the nature of proteins.

5.2.5 Other general features of the CreA protein:

There are four asparagine-X-serine/threonine motifs in the protein that are potential sites for N-linked glycosylation (Marshall 1972, Modrow and Wolf 1986) located at positions 18, 28, 191 and 283. However, it has been reported that the presence of a proline either between the asparagine and the serine/threonine or C-terminal to the serine/threonine as is the case with the predicted site at 191, will completely suppress N-glycosylation (Bause 1983). Therefore, the asparagine-X-serine motif at 191 is less likely to be a potential N-glycosylation site.

The phosphorylation/dephosphorylation of proteins is a common way in which the activities of proteins are regulated. Protein kinases alter the functions of their target proteins by
phosphorylating specific serine, threonine and tyrosine residues (reviewed by Kemp and Pearson 1990). In the absence of regulation of creA at the transcriptional level there remains the possibility that, with the high serine content, the CreA protein is activated/inactivated by specific phosphorylation. Numerous potential sites for phosphorylation in the CreA protein are identified using the PROSITE program (Bairoch 1990). These include consensus target sites for cAMP-dependent protein kinase, protein kinase C and caesin kinase II. The consensus sequence for cAMP-dependent protein kinase occurs between positions 101-105 in the amino acid sequence, placing it within the second zinc finger in the CreA protein. It would therefore be of interest to determine if this site is in fact a target for cAMP-dependent protein kinase and if so, what effect phosphorylation/dephosphorylation would have on the binding capacity of zinc finger II.

The Chou and Fasman (1978) and Garnier, Osguthorpe and Robson (1978) predictions as to the secondary structure of the putative CreA protein are shown in Figures 5.2.11 and 5.2.12.

5.3 SUMMARY:

2.7kb spanning the creA gene from A. nidulans has been sequenced. The derived amino acid sequence predicts that the creA gene encodes a protein of 390 amino acids which contains two zinc finger DNA binding motifs. The putative CreA protein is likely to be phosphorylated as it has a high serine content and this may represent a mechanism by which the activity of the protein is regulated. A repressor function is suggested for the CreA protein by the presence
FIGURE 5.2.11: The predicted secondary structures in the putative CreA protein according to Chou and Fasman (1978). Helices are shown with a sine wave, Beta sheets with a sharp saw-tooth wave, turns with 180 degree turns and coils with a dull saw-tooth wave. Hydrophilicity is superimposed over the wave and the size of these symbols is proportional to the value of these measures. Possible glycosylation sites are marked with O (Gribskov and Devereux 1986, Devereux 1984).
PLOTSTRUCTURE of: crea pep ck: 1381
TRANSLATE of: creag seq check: 4370 from: 827 to: 1088

Cheu-Fasman Prediction
November 20, 1990 13:25

KO Hydrophilicity >=1.3
KO Hydrophobicity >=1.3

Hydnophilicity >-1.3
Hydnophobicity >-1.3

COOH
FIGURE 5.2.12: The predicted secondary structures in the putative CreA protein according to Garnier et al. (1978). Helices are shown with a sine wave, Beta sheets with a sharp saw-tooth wave, turns with 180 degree turns and coils with a dull saw-tooth wave. Hydrophilicity is superimposed over the wave and the size of these symbols is proportional to the value of these measures. Possible glycosylation sites are marked with o (Gribskov and Devereux 1986, Devereux 1984).
of an alanine rich region which exists in other proteins where they have been identified as being domains conferring repressor function. That CreA may be a DNA binding repressor molecule is consistent with the genetical interpretation of creA being a negative regulator directly involved in carbon catabolite repression in A. nidulans. Experiments are now underway in this laboratory to confirm whether or not the CreA protein is able to bind to specific DNA sequences in the upstream regions of genes known to be regulated by the creA gene product.
CHAPTER SIX
CONCLUDING DISCUSSION.

The aims of the study outlined in this thesis were to isolate and characterize the wide domain regulatory gene, creA, from Aspergillus nidulans. This study was undertaken in order to begin to understand the molecular mechanisms of carbon catabolite repression in A. nidulans and more specifically to determine as far as possible how the product of the creA gene may be involved in this type of gene regulation.

A region containing the creA gene was cloned by complementation of the creA204 mutation using a plasmid library followed by the rescue of complementing sequences. This clone was subsequently used to isolate a larger genomic fragment containing the creA gene. The clone was also shown to complement several mutant alleles, including creA<sup>d</sup>-30 which has a translocation breakpoint within the creA coding sequence (Arst et al. 1990). The clone was shown to be homologous to a region which is disrupted in creA<sup>d</sup>-30 which allowed the inversion breakpoint to be mapped to the 250bp PstI fragment in the creA coding region.

Northern analyses using the cloned sequence as a probe demonstrated that creA is constitutively expressed at relatively low levels, expression levels only varying by about two fold depending on the available carbon source. Increased expression levels were found
in multicopy transformants and although these transformants grew as wildtype on a variety of carbon sources, there was evidence that increased creA copy number may lead to "tighter" carbon catabolite repression of at least some enzyme activities subject to control by the creA gene product. This could be investigated further by determining the effects of multiple copies of the creA gene under repressing and derepressing conditions on the mRNA levels for those genes, including alcR, alcA and amdS, which are subject to regulation by creA.

Strains containing the creA204, creA<sup>d</sup>-1, creA220, creA225 and creA<sup>d</sup>-30 mutations were all shown to produce creA mRNA, therefore suggesting that this group of alleles were not total lack of function creA mutants. A complete deletion of the creA coding region was constructed in a diploid strain using <i>in vitro</i> techniques. Analysis of this strain demonstrated that the phenotype for complete loss of function for creA is lethality. The creA mutant alleles isolated by various workers to date must therefore represent a class of mutants which have an altered function or lack some non-essential domain while retaining the essential function of the protein.

The physical analysis of the creA gene has included the determination of the nucleotide sequence of a 2.7kb region spanning the gene and the analysis of the structure of its transcript to determine the extent of the coding region. Analysis of the sequence showed an open reading frame from the first ATG in the transcribed region extending 1170bp to an inframe stop codon. The analysis of the effects of <i>in vitro</i> generated mutations or deletions is required to determine the functional significance of C-T rich sequences and other promoter-like elements in the upstream region of the creA gene. There
are no introns in the creA gene and the derived amino acid sequence predicts that the gene encodes a relatively small protein of 390 amino acids in size. The putative creA protein contains a number of features characteristic of DNA binding regulatory proteins including numerous S/TPXX motifs, two consecutive zinc finger DNA binding domains of the C₂H₂ class and an alanine rich region. The alanine rich region shows sequence similarity to domains in other proteins known to be important for transcriptional repression.

The mode of action of the creA gene product can be described by at least two models. One predicts that the creA gene encodes a DNA binding repressor protein involved in carbon catabolite repression, and that complete derepression of all the functions under creA control is lethal to the cell. Alternatively, the creA gene could encode a DNA binding repressor protein, but also has an essential activator function which as yet has not been genetically defined. Many eukaryotic transcriptional repressors, including the homeobox proteins and other zinc finger proteins, have been shown to contain both repressor and activator domains (reviewed by Levine and Manley 1989). Some of these proteins can function as either depending on the circumstances. For example, the relative contributions of these opposing domains could result from the conformation of the protein which can be influenced by several factors, including its affinity for DNA binding sites and interactions with other proteins. The comparison of activator domains from various transcription factors shows that although they are generally rich in acidic amino acids, there is a lack of a consensus sequence. It has been suggested that this low specificity could be sufficient if interactions between bound proteins are important for specific transcriptional activation
(reviewed by Guarente 1988). The finding that a complete deletion of the creA gene is lethal in a haploid strain supports the suggestion that there could be an essential activator domain, as well as a repressor domain, in the protein. It is interesting to speculate that the phenotypes of the creA mutant alleles may result from an impaired repressor function, while the lethality of a null allele may result from the loss of an essential activator function. Information from the sequences of creA mutants will determine which domain(s) in the CreA protein are required for repression, while the analysis of diploids transformed with in vitro generated mutant alleles may provide the necessary information required to determine whether an activator as well as a repressor domain is present in the protein.

There are a number of possible mechanisms by which CreA may act negatively to repress the transcription of regulatory and structural genes subject to carbon catabolite repression. These include repression mediated by the direct inhibition of binding of positively acting transcription factors, by the blocking of the activation of transcription or by inactivating the transcription initiation complex itself. Competition between the binding of negative regulators and particular positively acting transcription factors is commonly found in prokaryotic gene regulation (reviewed by Renkawitz 1990). This may involve the competition between negative regulators and general transcription factors (such as RNA polymerase I) at or near transcription start points or between more specific transcriptional activators upstream from the start point of transcription. In eukaryotes, it appears that repressors are commonly found to prevent the activation of transcription by positive factors which are bound to their target sites (reviewed by Levine and Manley 1989, reviewed
by Renkawitz 1990). Repression in this case involves the binding of activating and repressing proteins to adjacent non-overlapping DNA sequences and protein-protein interactions between the repressor and activator preventing the activator from making the correct contacts with the transcription complex. Alternatively, it is also feasible that CreA directly represses transcription by binding to defined sites, possibly even at a distance from the target promoters, and interferes with the formation or activity of the basal transcription complex leading to a reduced rate of transcription initiation (commonly called "silencing") (reviewed by Renkawitz 1990).

To further define how creA may act as a transcriptional repressor it is important to determine the position of putative binding sites in the upstream or promoter regions of its target genes. Many of the genes subject to control by carbon catabolite repression and the product of the creA gene have been cloned. It is therefore of interest to identify upstream sequence elements in these genes which are necessary and sufficient for creA mediated regulation. It is not known whether carbon repression acts via regulatory genes or directly on the 5' regions of genes encoding enzymes, or on both as appears to be the case for the genes involved in ethanol metabolism in A. nidulans (Felenbok et al. 1989). Features of the derived amino acid sequence of the creA polypeptide indicate that it is likely to be a DNA binding protein. Therefore, binding of the protein to putative DNA target sites could be investigated by mobility shift assays and footprinting studies. Although such experiments are required to confirm that creA is a DNA binding protein, they will also serve to locate the positions of such sites and conditions under which binding occurs.
Different proteins containing similar structural motifs such as a zinc finger are able to recognize and bind to a variety of DNA sequences in different target genes, although probably also with different affinities. Investigations into the binding specificity and affinities for different target sites of *in vitro* generated mutations in the zinc finger regions of the CreA protein and their binding sites could be carried out. This would allow the characterization of important residues in the fingers required for their specificity, the relative importance of both the fingers for CreA function and/or binding and the critical nucleotides in their target binding sites.

Titration analysis is another valuable approach for studying gene regulation. This involves the construction of strains carrying multiple copies of a target site for a *trans*-acting regulatory protein which is present in cells in limiting amounts. It results in the titration of the regulatory molecule such that reduced levels are insufficient for affecting complete regulation of its target genes. Titration studies in *Aspergillus* have provided valuable information on the relative affinities of different binding sites for *trans*-acting regulatory proteins, the effects of mutations in regulatory genes on the binding capacity of the protein and on the relative amounts of regulatory proteins present in certain strains (Andrianopoulos and Hynes 1988, Hynes and Kelly 1987, Hynes et al. 1988). Although this approach to studying gene regulation in *Aspergillus* has only involved the titration of positively acting regulatory proteins such as *amdR* and *facB*, similar experiments should be possible to show the titration of transcriptional repressors, thus defining target sites and relative affinities to such sites. If CreA is present in limiting amounts in cells and does in fact have a
trans-acting repressor function, transformants containing multiple copies of its target sites should lead to overexpression of the regulatory and/or structural genes subject to direct regulation by creA in the presence of glucose. Furthermore, anti-titration should be possible by increasing creA copy number and expression in these strains.

The creA clone from A. nidulans was shown to hybridize to sequences from a number of other fungi. It will be of interest to isolate and characterize these sequences to determine if a common mechanism of carbon catabolite repression exists in these species. This would also allow comparisons to the A. nidulans sequence to identify conserved regions which may define further the important regions for creA function. It would also be of interest to investigate whether the deletion of such sequences from these other species is lethal and to determine if these genes are able to confer creA-related functions in heterologous expression systems.

Although this study has concentrated on the molecular analysis of the creA gene, other loci have been genetically identified as being involved in carbon catabolite repression in A. nidulans (Hynes and Kelly 1977, Kelly and Hynes 1977). Therefore, the isolation and molecular characterization of creB and creC, in particular, are desirable in order to more fully understand the roles of these genes in carbon repression. Attempts to isolate these genes using the same techniques used to isolate creA are underway in this laboratory.

The control of gene expression by trans-acting regulatory genes such as creA requires that the regulatory protein has a number of functional properties. These include, a capacity to enter and
accumulate in the nucleus, an ability to bind to specific target cis-acting elements and/or to interact with other products that bind to these sites leading to up or down regulation mediated by these other molecules and following this, to activate or repress target gene expression either directly or indirectly. Such transcription factors must also respond accordingly by some effect on one of these activities, to specific environmental stimuli.

The identification of important domains using \textit{in vitro} generated mutations and deletions and \textit{in vivo} assays for the function of regulatory gene products has been achieved for many regulatory proteins including GCN4 (Hope and Struhl 1986), GAL4 (Johnson \textit{et al.} 1986), ADRI (Blumberg \textit{et al.} 1987) and NIT2 (Fu and Marzluf 1990). It has become apparent from some of these studies that a common property of regulatory proteins is that relatively large regions of these proteins may be deleted without noticeably altering their function. For example, mutational analyses of the positively acting NIT2 regulatory protein from \textit{N. crassa} have demonstrated that while both the zinc finger domains and a downstream basic region are critical for DNA binding and its transcriptional activation function, deletion of 21\% of the COOH-terminus of the protein maintained full activity (Fu and Marzluf 1990). Similarly, although analysis of the putative CreA amino acid sequence has identified at least two domains in the protein, one a DNA binding region and another possibly required for repressor function, the molecular dissection of the CreA protein should allow the functional domains and the important regions required for CreA activity to be identified.

The studies reported in this thesis provide a solid basis for further investigations into the precise nature and mechanisms
involved in creA mediated gene regulation as a model for global gene regulation in eukaryotes. Future analyses of the creA gene product should clarify the function of the protein at the molecular level.
REFERENCES.


Structure of the cro repressor from bacteriophage lambda and its
interaction with DNA. Nature 290: 754-758.
positively acting regulatory gene amdR from Aspergillus nidulans.
Arst, H.N. 1977. Some genetical aspects of ornithine metabolism in
in Aspergillus nidulans. pp. 131-146. In J.E. Smith and J.A.
Pateman (eds.) Genetics and physiology of Aspergillus. Academic
Press, Inc. N.Y.
nidulans with an internally located cis-acting regulatory region.
Arst, H.N. and Scazzocchio, C. 1975. Initiator constitutive mutation
with an "up"-promoter effect in Aspergillus nidulans. Nature 254:
31-34.
Arst, H.N. and Scazzocchio, C. 1985. Formal genetics and molecular
biology of the control of gene expression in Aspergillus
associated loss-of-function mutation in the nitrogen metabolite
repression regulatory gene of Aspergillus nidulans can revert


Goto, T., Macdonald, P. and Maniatis, T. 1989. Early and late periodic patterns of evenskipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell 57: 413-422.


