Molecular characterisation of mutations in the *qutA* activator gene for quinic acid utilisation in *Aspergillus nidulans*

by Ian Levesley

B.Sc (Hons.) Genetics

University of Newcastle Upon Tyne

Thesis submitted for the degree of Ph. D. in the Department of Genetics, University of Leicester.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Molecular characterisation of mutations in the *qutA* activator gene for quinic acid utilisation in *Aspergillus nidulans*

by Ian Levesley

Aspergillus nidulans utilises quinic acid as a sole carbon source for growth. The inducible quinic acid utilisation (*qut)* pathway shares intermediates with the constitutive biosynthetic shikimate pathway, which is maintained despite competition from the quinate pathway enzymes. Genetic analysis isolated a large number of non-quinic acid utilising mutants, including pleiotropic regulatory mutants in the *qutA* activator and *qutR* repressor genes. The entire 19kb *qut* gene cluster on chromosome VIII has been cloned and sequenced, as has the *aromA* gene encoding a penta-functional polypeptide from the shikimate pathway. Molecular analysis suggests that the *qutA* and *qutR* genes evolved by duplication and cleavage of the *aromA* gene (Hawkins *et al.*, 1993), suggesting possible molecular roles. However, a precise molecular mechanism for the function of the regulatory proteins has yet to be elucidated.

A contribution to this analysis is detailed in this thesis, extending the identification of functional motifs in the *qutA* activator. A genetic map of non-inducible *qutA* mutant strains has been constructed. Mutants yield either recessive, semi-dominant or dominant phenotypes in diploid strains with the wild type activator. Oligonucleotide primers were designed to amplify specific regions of the *qutA* gene, by PCR, from wild type and mutant genomic DNA. SSCP analysis of the PCR amplified DNA was used to align the genetic and physical maps. PCR amplified DNA was subsequently used in sequencing reactions to characterise seventeen mutations at the nucleotide level and infer changes in the protein.

The genetic and physical maps proved to be co-linear, non-inducible *qutA* mutations mapped to the 3' half of the gene, implying that mutations in the 5' half of the gene produce a transcriptionally active protein, indicating a centrally located transcriptional activation motif. The two dominant non-inducible mutations mapped to the junction of the two putative domains of the protein predicted by its homology to *aromA*, strongly supporting the proposed bi-domain structure. The results presented in this thesis together with other recent work discussed enable a revision of the molecular model, and provide a strong foundation for the *in vitro* study of the two regulatory proteins.

Table of contents

Chapter Two
Methods and Materials

Chapter Three

Preliminary mapping of *qutA* mutations in *A.nidulans*

Chapter Four

Part I

Approaches to sequencing regions of the *qutA* gene

Part II

Analysis of the putative Region 1 clones

Chapter Five

DNA sequence analysis of *qutA* Region 1

Chapter Six

DNA sequence analysis of Region 2 of the *qutA* gene

Chapter Seven

DNA sequence analysis of Region 3 of the *qutA* gene

Chapter Eight
Discussion

Abbreviations

One and three letter amino acid abbreviations

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Glu	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Nucleic acid Abreviation

DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
ssDNA	single stranded DNA
ATP	adenosine triphosphate
NTP	nucleotide triphosphate
A (dATP)	deoxyadenosine triphosphate
C (dCTP)	deoxycytidine triphosphate
G (dGTP)	deoxyguanosine triphosphate
T (dTTP)	deoxythymidine triphosphate
N (dNTP)	deoxynucleoside triphosphate
ddATP	dideoxyadenosine triphosphate
ddCTP	dideoxycytidine triphosphate
ddGTP	dideoxyguanosine triphosphate
ddTTP	dideoxythymidine triphosphate
DNase	Deoxyribonuclease
RNA	Ribonucleic acid
RNase	ribonuclease
bp	base pairs
kb	kilobase
nt	nucleotides

Other abbreviations

aa	amino acid
Amp^r	Ampicillin resistance
aromA	Gene encoding a pentafunctional enzyme of the shikimate pathway
AROM	pentafunctional shikimate pathway enzyme
BSA	Bovine serum albumin
BBL	Baltimore biological laboratories
D	Dominant allele of *qutA*
DHQ	Dehydroquinate
DHQS	Dehdroquinate synthase
DHS	Dehydroshikimate
DMSO	Dimethyl sulphoxide
DQ	Dehydroquniase
DTT	Dithiothreitol
EDTA	Sodium ethylene-diamine-tetra-acetic acid
EMBL	European Molecular Biology Laboratory
EPSP	Enol pyruvate shikimate phosphate
EPSPS	Enol pyruvate shikimate phosphate synthase
EtBr	Ethidium Bromide
hr	hour(s)
IMS	Industrial methylated spirit
IPTG	Isopropyl-β-D-galactopyranoside
LA	Luria agar
LB	Luria broth
MA	Minimal agar
MEA	Malt extract agar
MM	Minimal media
min	minute(s)
Nal^r	Naladixic acid resistance
ONPG	Ortho-nitryl-phenyl-galactoside
ORF	Open reading frame
PABA	Para amino benzoic acid
PCA	Protocatechuic acid
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
PYRO	Pyrodoxin
QA	Quinic acid
qut	Quinate utilisation (*A.nidulans*)

qutA	*qutA* gene
QUTA	protein produced by the *qutA* gene
R	recessive allele of *qutA*
RF	recombination frequency
SD	semi-dominant allele of *qutA*
SDS	Sodium dodecyl sulphate
SDH	Shikimate dehydrogenase
sec	second(s)
SK	Shikimate kinase
SSCP	Single Stranded Conformational Polymorphism
Ta	annealing temperature
TAE	Tris-acetate EDTA buffer
TBE	Tris-borate EDTA buffer
TE	Tris-EDTA
TEMED	N, N, N', N'-tetramethylethylenediamide
Tm	melting temperature
Tris	Tris (hydroxymethyl methylamine)
	[2-amino-2-(hydroxymethyl) propane-1,3-diol, (tris)]
URS	Upstream repression sequence
UAS	Upstream activation sequence
UV	Ultraviolet
v/v	volume per volume
wt	wild-type
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Chapter One

Gene regulation in lower eukaryotes

Gene expression in lower eukaryotes is regulated at three levels: transcriptional, post-transcriptional and post-translational regulation. Transcriptional control is the predominant form of gene regulation. Expression or protein function may have complex forms of post-translational or post-transcriptional control in addition to transcriptional control notably in the case of some regulatory genes. This introduction centres upon the structure and function of transcriptional regulatory proteins. The brief discussion of other regulatory mechanisms below concentrates upon how two comprehensively studied transcriptional activators GCN4 and GAL4 are themselves subject to post-translational and post-transcriptional regulation respectively.

Post-transcriptional regulation occurs in eukaryotic mRNA processing activities such as alternate intron splicing (Review, Leff *et al.*, 1986), regulation of RNA transport (Malim *et al.*, 1989), mRNA stability (Marzluff & Pandey, 1988; Review, Raghow, 1987) and translational regulation. The yeast regulatory protein GCN4 is a transcriptional activator involved in general amino acid control, the response to amino acid starvation (Chapter 1.1). The protein activates expression of more than 30 genes in 9 different biosynthetic pathways (Hope & Struhl, 1985). The GCN4 mRNA encodes a 600bp leader sequence prior to the activator protein translational start codon (most yeast mRNAs have a 60-90bp leader sequence {Gurr *et al.*, 1987}). GCN4 mRNA levels remain constant in all conditions. Regulation of protein levels occurs via four short open reading frames (ORFs) in the mRNA leader sequence. The proteins GCN1, 2 and 3, increase levels of the GCN4 protein. GCD1, 10 and 13 decrease protein levels in non-starvation conditions (Hinnesbusch, 1985; Mueller *et al.*, 1987; Review, Hinnesbusch, 1988).

Post-translational regulation occurs by three mechanisms: ligand binding, protein-protein interaction and protein modification. The yeast transcriptional activator ACE1 regulates expression of metallothionine in response to copper. Binding of copper radically alters activator protein structure allowing the promoter sequence upstream of the target gene to be recognised (Furst *et al.*, 1988). GCN4 translation is dependant upon several positive and negatively acting factors, including a positively acting factor GCN2. GCN2 is a protein kinase that affects re-initiation of yeast ribosomes (Tzamarias & Thrieos, 1988) and therefore GCN4 expression by phosphorylation (protein modification) of one of the other GCN or GCD (Roussou *et al.*, 1988). Binding of uncharged tRNA

molecules activates GCN2 protein kinase function (Wek *et al.*, 1989). Thus, a protein modification event, catalysed by GCN2, itself regulated by a ligand binding event, controls post-transcriptional regulation of GCN4. The yeast GAL regulatory system (Chapter 1.2) provides incidence of all three forms of post-translational regulation. The GAL4 protein is a positively acting transcriptional regulator and GAL80, a specific repressor, represses GAL4 by protein-protein interaction (Leuther & Johnston, 1991). Ligand binding by GAL4 depresses GAL4 activity. The ligand is thought to be a small metabolite formed as a consequence of galactose metabolism by the *gal3* gene product (Broach, 1979). Additionally, a series of specific phosphorylation events apparently enhances the activity of the de-repressed GAL4 activator (Mylin *et al.*, 1990).

Transcriptional regulation is mediated by both general and specific trans-acting proteins that bind to DNA control elements in the promoter region of target genes. The binding of transcriptional regulators to their target sequence requires the action of chaperone transcription factors (Cote *et al.*, 1994). Once bound to their target sequences the trans-acting regulators interact with further factors to promote or repress transcription of the target gene.

Two classes of cis-acting DNA sequences or promoter elements mediate gene transcription in eukaryotes. The first class, the core promoter, contains the binding site for the native TATA factor (the primary DNA binding factors in the formation of the transcription complex), and controls the location of transcription initiation. The core promoter alone seems to be sufficient for basal level expression of some genes. The second class of promoter element is the binding site for regulatory factors that determine the rate at which RNA polymerase II initiates new rounds of transcription from the core promoter. Activator proteins act either directly with components of the native TATA factor or indirectly through mediator or adapter factors. The activator causes signal transmission to the carboxyl terminal domain of RNA polymerase II where it interacts with the native TATA factor (Colgan *et al.*, 1993; Yankulov *et al.*, 1994). The interaction affects the integration of general transcription factors into the pre-initiation complex and its stabilisation to form the transcription elongation complex (Choy & Green, 1993).

General transcription factors are essential for gene expression and therefore growth. Low levels of these proteins in eukaryotes and the instability of the transcription complex have made study of the process of transcription initiation difficult. However, recent advances in techniques for the cloning, expression and detection of gene products have assisted the purification of the factors involved. *In vitro* techniques such as template competition, restriction site protection, 'foot-printing' and electrophoretic mobility shift assays have allowed provided a picture of the factors involved and the order and kinetics of their assembly. This larger subject is beyond the scope of the present

discussion on transcription regulators but has recently been reviewed by Conaway & Conaway, (1993).

Transcriptional activation of gene expression is always mediated by the process described above. Different mechanisms affect transcription repression. A repressor may bind DNA competitively with an activator. Less commonly, the repressor may bind DNA and affect regulation of a whole region of the genome such as at the *HMRa* mating locus in *Saccharomyces cerevisiae* (Loo & Rine, 1994). Alternatively, a repressor protein that interacts directly with a transcriptional activator preventing the formation of an active transcription complex.

Genetic analysis identified regulatory genes in lower as pleiotropic mutations affecting more than one function in the cell; these regulatory genes fell into two phenotypic classes. Different mutant alleles at a single locus that alter functions in more than one pathway characterised the first class. They were complex in that some mutant alleles repressed function in one pathway, de-repressed another whilst remaining normally regulated for a third. The diverse effects upon genes of different pathways led to these genes being termed 'wide domain' regulators. Mutant alleles of the second class of regulatory affect, either negatively or positively, all the functions of a single pathway and are 'narrow domain' regulators.

Transcriptional regulatory genes and their products

The next section will outline some approaches used in the study of regulatory protein structure and function, thus identifying methods of studying regulatory proteins in general and the *qutA* activator in particular. The examples are selected to give information about methods of studying or data obtained about transcriptional activators, rather than the systems they regulate. A 'motif' typically refers to a stretch of nucleotides of a specific sequence such as a promoter element or motif. Here this term includes short amino acid sequences such as the peptide encoding a DNA binding function. The transcriptional regulators described below have been studied by the comparison and characterisation of these functional motifs. These homologous functional motifs appear within a backbone structure that show little or no homology between transcriptional regulators. The origin and possible functions for some backbone sequences are described below. The term 'Domain structure' describes this more global view of regulatory proteins. Domain, therefore, refers to a larger region of a protein where three dimensional structure, rather than sequence, defines domain boundaries. Thus, a protein may have one or more domain each of which may contain a number of motifs.

1.1 Wide domain regulatory genes and their products

This section describes regulatory genes involved in ensuring an integrated physiological response to different aspects of general cellular metabolism. Nitrogen and carbon utilisation in *Aspergillus nidulans* represent catabolic processes. Similar systems exist in *A.nidulans* for the utilisation of phosphorus (Arst & Scazzochio, 1985) and sulphur (Caddick *et al.*, 1986b) nutritional sources. The amino acid starvation response represents biosynthetic processes. Amino acid availability balances induction of pathways for amino acid synthesis ensuring efficient use of cellular resources. The section concludes with an example of how different promoter elements in a single gene confer global regulation upon its expression and a summary of the findings.

1.1.1 Nitrogen metabolite repression

Several pathways exist for the utilisation of different nitrogen sources. Glutamine and ammonium groups use is preferential to more energy demanding nitrogen sources such as nitrates. The *areA* gene of *A.nidulans* encodes a transcriptional activator that promotes expression of pathways for the less efficiently utilised nitrogen sources (Arst & Cove, 1973). Glutamine and ammonium (probably via conversion to glutamine) inactivate the *areA* gene product. This ensures that enzymes required for utilisation of an energy demanding substrate are only expressed in the absence of a preferred nitrogen source (Arst & MacDonald, 1973: Kinghorn & Pateman, 1973: Pateman *et al.*, 1973: Hynes, 1974: Arst *et al.*, 1982: Cornwell & MacDonald, 1984).

Deletion analysis: The wild-type *A.nidulans areA* gene encodes a 719 aa protein (Caddick *et al.*, 1986a). Manipulation of the cloned *areA* gene *in vitro* produced a number of partially deleted clones used to transform *areA*$^-$ mutant strains of *A.nidulans*. Deletion analysis and analysis of translocation mutations with the break points in the *areA* locus, indicate that the essential region of the protein lies between residues 343-595 (Arst *et al.*, 1989: Kudla *et al.*, 1990).

Sequence motifs: This region contains two putative motifs with transcription functions, a Zinc finger DNA binding motif and a highly acidic transcriptional activation motif. Sequence analysis of the *areA102* allele revealed a single amino acid substitution of leucine to valine. The substitution lies in the central portion of the zinc finger loop. The *areA30* and *31* alleles, obtained by reversion of *areA102* strains, have a reciprocal phenotype to *areA102*. They show increased growth on formamide and xanthine as nitrogen sources and decreased levels of acetamidase, histidase and citrulline permease (Arst & Scazzochio, 1975; 1985). DNA sequence analysis of the *areA30* and *31* alleles revealed they were identical mutations. Methionine substitutes valine at the same amino acid altered in the *areA102* allele. This implicates the central portion of the Zinc finger motif in promoter element recognition (Kudla *et al.*, 1990).

Domain identity: A recent proposal suggests that the AREA protein has a tri-domain structure. The three domains being G (aa1-251), C (aa252-508) and F (aa509-720), evolved from the tri-domain β subunit of the anthranilite synthase AS complex (Walker & De Moss, 1983; 1986). Acquisition of a zinc finger DNA binding motif altered the F domain. The G domain retains the glutamine binding function encoded partially by the β subunit and partially by the α subunit of the AS complex. The proposal further suggested that the *tamA* (Kinghorn & Pateman, 1975) or the *meaB* gene (Arst, 1990) encoded a protein derived from the α subunit of the AS complex. In this model glutamine and the *meaB* or *tamA* products interact with AREA causing repression (Hawkins *et al.*, 1994). The predictions for protein structure and function made by these proposals are currently being investigated.

1.1.2 **Carbon catabolite repression**

Identification of genes involved in carbon catabolite repression in *A.nidulans* used strains containing recessive mutant alleles of *areA*. Acetamidine and L-proline serve as both nitrogen and carbon sources, and the enzymes required for utilisation are under nitrogen and carbon catabolite repression. Mutant *areA* strains that utilise acetamidine and L-proline in the presence of glutamine or ammonium, but not in the presence of glucose, were selected. Following UV mutagenesis, strains capable of utilising acetamidine or L-proline in the presence of glucose were retained. Screening of the retained strains identified and rejected those deficient in the uptake or utilisation of glucose. Genetic analysis identified second site mutations mapping to three unlinked loci designated *creA, creB* and *creC* (Bailey & Arst, 1975; Hynes & Kelly, 1977). The *creA* gene was proposed to encode a transcriptional repressor (Bailey & Arst, 1975).

Sequence motifs: The *creA* gene was cloned from *A.nidulans* and sequenced (Dowzer & Kelly, 1989). The gene produces a 415 amino acid polypeptide (CREA) encoding a putative zinc finger (C_2H_2) and alanine rich region indicative of a DNA binding repressor (Dowzer & Kelly, 1991).

Physical studies: CREA binds to a hexa-nucleotide promoter sequence in the physiologically related target genes. At some promoters the protein binds to multiple sites around the consensus sequence co-operatively stabilising the repression (Espeso & Penalva, 1994). Catabolite repression of the proline gene cluster in *A.nidulans* by CREA is dependent upon two different hexa-nucleotide binding sites separated by a single base (Cubero & Scazzochio, 1994). Competitive binding of the *CREA* and *ALCR* to overlapping promoter sites mediates repression of the autoregulatory *alcR* gene (encoding a transcriptional activator for the ethanol utilisation genes *alcA* and *aldA*). *alcA* and *aldA,* also posses the CREA binding sites in their own promoters (Mathieu & Felenbok, 1994). The C-terminal 145aa of CREA are sufficient for repression. CREA is autoregulated with binding sites approximately 400 and 500bp upstream of the translation start. The fact that CREA levels are higher

in non-repressing conditions suggests that CREA has some affect under conditions not generally considered carbon catabolite repressing (Shroff *et al.*, 1996)

Carbon catabolite repression of quinate utilisation (*qut*): Analysis of the *qut* cluster promoter regions for CREA binding sites reveales the following finding. Within the entire cluster only one good match (11 out of 13 bases) with the Cubero & Scazzochio (1994) consensus sequence was identified. This match was upstream of the *qutD* permease encoding gene (This thesis). No CREA mutants affecting quinate utilisation have been identified. However, these were looking for altered enzyme levels in the induced state (Hynes & Kelly, 1977) and not specifically for quinate uptake. Little is known of the structure or function of the *creB* and *creC* alleles and their products. Reduced quinate dehydrogenase (*qutB*) activity is observed with some mutant alleles of *creB* and *creC,* but induction remains sensitive to carbon catabolite repression.

1.1.3 **Amino acid starvation**

The response to this situation is known as cross pathway regulation in *Neurospora crassa* (Cariotis *et al.,* 1974a: Cariotis & Jones, 1974b) and *A.nidulans* (Piotrowaska, 1980) and general aa control in *S.cerevisiae* (Niederberger *et al.*, 1981).

In *S.cerevisiae* pathways for the synthesis of isoleucine/valine, histidine, lysine and arginine are increased 2-10 fold (at transcription level), if any one of these amino acids becomes depleted (Hope & Struhl, 1985: Reviewed by Hinnebusch, 1988). DNA sequence analysis of the promoter region of genes subject to general aa control identified a common promoter element necessary for their induction. Strains containing a mutation in the *GCN4* gene failed to induce gene expression in response to amino acid starvation (Hinnebusch & Fink, 1983: Penn *et al.,* 1983). All known mutations in other trans-acting genes affected levels of GCN4 protein (Mueller & Hinnebusch, 1986). GCN1, 2 and 3 increase GCN4 levels whilst GCD1, 10 and 13 repress expression in non-starvation conditions (Hinnebusch, 1985: Mueller *et al.*, 1987).

Physical studies: *In vivo* and *in vitro* DNA binding experiments showed that the protein binds to the common promoter element upstream of the regulated genes (Hope & Struhl, 1985: Ardnt & Fink, 1986).

Sequence motifs: GCN4 contains a DNA binding motif and a large acidic transcriptional activation motif (Hope & Struhl, 1986). The DNA binding motif forms a leucine zipper. Leucine residues in a helical structure formed by each monomeric protein inter-digitate to form a dimer. The dimerised interface binds non-specifically to DNA, and the basic residues immediately adjacent in the amino terminal direction provide the specificity of binding (Agre *et al.,* 1989).

Deletion analysis: Deletion analysis located a transcriptional activation motif, in which as few as 35-40 aa are sufficient for wild-type levels of transcriptional activation (Hope *et al.,* 1988).

Portions of the acidic motif with different primary sequences are equally efficient transcriptional activators. Apparent correlation between the length of the acidic region and levels of transcriptional activation suggested that the different acidic regions may act additively (Struhl, 1989). Wek *et al.*, 1989 determined the relationship between GCN4 function and amino acid availability. Protein kinase activity plays a key role in general amino acid control.

GCN2 and Domain identity: The *GCN2* gene product posses or regulates protein kinase activity *in vitro* and the protein showed identity to a particular class of protein kinases (Tzamarias & Thrieos, 1988: Roussou *et al.*, 1988). Re-examination of the *GCN2* nucleotide sequence identified an extension to the ORF recognised originally. This extended region showed no protein identity with the protein kinases. At low amino acid concentrations, the ratio of tRNA bound to their respective amino acids compared to uncharged tRNA decreases. The extended region of *GCN2* sequence encodes a polypeptide with 22% identity at the amino acid level to the *S.cerevisiae*, *Escherichia coli* and human histidyl tRNA synthase (HisRS). Despite the low level identity the domain may recognise uncharged His-tRNA molecules and other uncharged tRNAs. GCN2 forms a bi-domain structure. The HisRS like domain binds uncharged tRNA molecules (a gauge for amino acid levels) that activates the protein kinase activity of the adjacent domain. One of the *GCD* or the other *GCN* products may be the substrate for the kinase activity. The bi-domain theory provides the link between known protein kinase activity in GCN4 regulation and amino acid availability (Wek *et al.*, 1989).

1.1.4 Cis-acting determinants of wide domain regulation

So far the discussion on wide domain control has centred upon the activator and repressor proteins that regulate expression by binding to promoter elements upstream of the gene. However, the presence of more than one class of promoter element preceding a gene can confer global regulation. The promoter region of the *amdS* gene, encoding acetamidase in *A.nidulans,* contains binding sites for three narrow domain transcriptional activators encoded by the *amdR*, *amdA* and *facB* genes. Additionally, two wide domain regulatory proteins bind the promoter region AREA, and CREA product (Hynes *et al.*, 1988: Adrionopolous & Hynes, 1990). The *amdR* product that responds to ω-amino acids induces *amdS* expression. Induction by the FACB and AMDA forms the major and minor forms of acetate induction respectively (Davis *et al.,* 1993). The AREA and CREA products are responsible for nitrogen and carbon catabolite control of *amdS* expression. Different regulatory proteins have a cumulative effect on expression of the *amdS* gene (Hynes *et al.*, 1989).

1.1.5 General features of wide domain regulation

Expression of physiologically related genes are negatively or positively affected by wide domain regulation. The regulatory proteins control expression by binding to promoter elements upstream of the genes they regulate. Different promoter elements in a gene ensure that a range of regulatory signals control the overall expression of each individual target gene.

1.2 Narrow domain regulatory genes and their products

Regulation of a single metabolic pathway involves a smaller number of regulatory proteins, additional transcription factors and regulated genes and is, therefore, more accessible for molecular analysis. Narrow domain transcriptional activators bind to promoter elements upstream of the genes they regulate. However, the activator becomes active in the presence of a small metabolite, usually a substrate or product of the pathway it regulates. Narrow domain regulatory proteins are commonly subject to one of two basic methods of post-translational regulation. Ligand binding, an activator binds an inducer potentiating its activity. Protein-protein interaction with a specific repressor binding may prevent activator function. The repressor binds the inducer, altering the conformation of the activator-repressor complex, allowing transcriptional activation.

1.2.1 Proline utilisation (*Put*) in *Saccharomyces cerevisiae*

S.cerevisiae utilises proline as a nitrogen source, when preferred nitrogen sources are unavailable. The enzymes encoded by the genes *Put1* (proline oxidase) and *Put2* (Δ-pyrroline-5-carboxylate dehydrogenase) convert proline to glutamate (Brandriss & Magasanik, 1979). Induced and basal level transcription of the *Put1* and *Put2* genes is regulated by the activator protein encoded by *Put3* (Brandriss, 1987).

Physical studies: *Put3* protein levels are similar in the presence or absence of proline, and the activator protein binds DNA in both conditions (Siddiqui & Brandriss, 1989: Axelrod *et al.*, 1991).

Sequence motifs: Cloning and sequencing the *Put3* gene revealed a 979 amino acid protein containing a Zinc bi-nuclear cluster motif and two putative acidic activation motifs (Marczak & Brandriss, 1989). *Put3* protein may bind proline causing a conformational change in the protein that allows contact with the transcription complex (Marczak & Brandriss, 1991).

1.2.2 **Ethanol utilisation in *Aspergillus nidulans***

The ALCR protein regulates Alcohol dehydrogenase (*alcA*) and aldehyde dehydrogenase (*aldR*) genes (Pateman *et al.,* 1983) transcription. Ethanol utilisation is subject to carbon catabolite repression mediated by the *creA* gene (Lockington *et al.*, 1987) via promoter motifs upstream of *alcR*, *aldR* and *alcA*.

Sequence Motifs:- DNA sequence analysis of *alcR* identified two overlapping out-of-phase ORFs each producing a 96kd protein (Felenbok *et al.,* 1988). Later work by Kulmberg *et al.*, (1991) showed that one of these ORF encodes a Zinc bi-nuclear cluster DNA binding motif.

Physical studies: Expression of the motif in *E.coli* produced a truncated protein that binds to a promoter element motif upstream of the *alcR* gene (Kulmberg *et al.,* 1992). Sequeval & Felenbok, (1994) have shown that the ALCR requires Zinc for function.

Domain identity: Hawkins *et al.,* (1994) suggest that the *alcR* activator evolved by fusion of the *alcA* and *aldR* genes. This proposal suggests that the transcriptional activator that retains the binding sites for alcohol or acetaldehyde or both, these form the inducing molecule that potentiates its activity. The activator itself binds the inducing molecule to become active in transcription negating the requirement for a separate repressor protein.

1.2.3 **Leucine biosynthesis (LEU) in S. cerevisiae**

The transcriptional activator *LEU3* (Fridden & Schimmel, 1988) regulates *LEU1, 2, 4* and *ILV2* involved in branched chain amino acid biosynthesis, and the *GDH1* gene involved in nitrogen metabolism. Expression of *LEU3* itself is subject to general amino acid control (Zhou *et al.*, 1987). The activator function depends entirely on the presence of the metabolite α-isopropylmalate (αIPM) an early intermediate in leucine biosynthesis (Sze *et al.*, 1992). In the absence of αIPM, *LEU3* protein acts as a repressor reducing expression of the regulated genes below the basal level expression observed in a strain carrying a *LEU3* deletion (Brisco & Kohlaw, 1990).

Sequence motifs: The *LEU3* activator gene has been cloned and sequenced. This identified two functional motifs (Zhou *et al.*, 1990): a zinc bi-nuclear cluster DNA binding motif (Bai & Kohlaw, 1991) and a transcriptional activation motif (Zhou & Kohlaw, 1990).

Physical studies: In the absence of the inducer, αIPM, the *LEU3* activator protein undergoes a conformational change. This prevents transcriptional activation and may affect the binding of the transcription complex to the DNA (Zhou *et al.*, 1993).

1.2.4 **Galactose utilisation (*GAL*) in *S.cerevisiae***

Douglas & Condi, (1953) reported upon the genetic control of galactose utilisation in *Saccharomyces*, Douglas & Pelroy, (1963) identified a gene regulating induction of the pathway.

The Douglas & Hawthorne, (1966) model for galactose regulation in *S.cerevisiae* serves as a paradigm for gene regulation in lower eukaryotes. The particular facility of *S.cerevisiae* for genetic manipulation and subsequently for molecular studies has made and maintained *GAL* regulation as one of the most comprehensively studied eukaryotic regulatory systems. The GAL4 gene encodes an 881 amino acid transcriptional activator protein (Laughon & Gesteland, 1984). GAL4 induces expression of the three genes required for galactose utilisation (Hopper & Rowe, 1978) upon binding to target promoter elements (Johnston & Davis, 1984).

Sequence motifs: The promoter elements have a 16bp palindromic consensus sequence (Giniger *et al.*, 1985: Bram & Kornberg, 1985). The amino terminus of the GAL4 protein encodes: a Zinc $(II)_2$ $Cysteine_6$ bi-nuclear cluster DNA binding motif (Pan & Coleman, 1990) and a nuclear localisation signal within the first 74 amino acids (Silver *et al.*, 1984). However, the nuclear localisation signal and Zinc cluster motif (aa 10 to 51) appear to be functionally separate (Silver *et al.*, 1986). The activator has two different peptide motifs capable of activating transcription (Ma & Ptashne, 1987a), proposed to form an α-helical structure. Of the two, it is the motif at the carboxyl terminus that is the functional element in the intact activator; this is also the site for activator-repressor protein binding (Ma & Ptashne, 1987b).

Galactose utilisation is under carbon catabolite repression (Reviews, Lamphier & Ptashne, 1992: Johnston *et al.*, 1994). The *GAL80* gene product binding blocks the function of the activator protein (Nogi & Fukasawa, 1984). In the presence of galactose the repressor binding alters conformation and the activator protein becomes transcriptionally active (Leuther & Johnston, 1991).

The actual metabolic signal for induction of the galactose system is unknown but involves an inducer or co-inducer produced by the *gal3* gene product (Torchia & Hopper, 1986). *gal3* mutants adapt slowly to the presence of intracellular galactose (Speigleman *et al.*, 1950: Rotman & Speigleman, 1953: Adams & Dalbec, 1977). Double mutants of *gal3* with *gal1, gal7* or *gal10* never become induced suggests that the inducer is the end product of galactose metabolism (UDP-glucose) modified by the *gal3* product (Broach, 1979). This mechanism ensures that the pathway is only induced if galactose is present and then only if the organism can utilise galactose.

1.2.5 **Quinate utilisation in *A.nidulans* and *N.crassa***

A transcriptional activator *qutA* (*qa-1F* in *N.crassa*) and a specific repressor *qutR* (*qa-1S* in *N.crassa*) regulate quinate utilisation in *A.nidulans* (Hawkins *et al.*, 1982) and *N.crassa* (Case & Giles, 1975). The quinic acid utilisation (*qut*) system of *A.nidulans* and particularly the *qutA* activator are the subjects of the remainder of this chapter and of the work in the thesis.

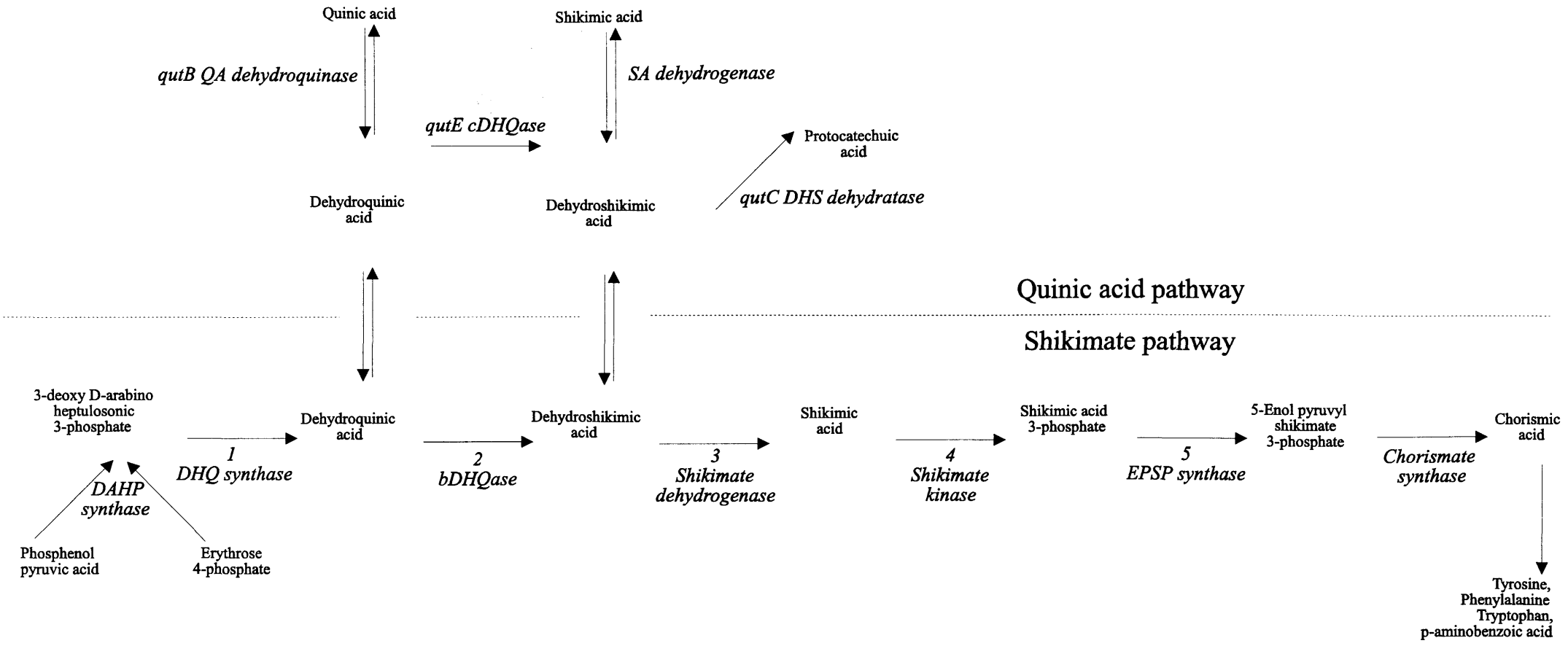
1.3 The Quinic acid utilisation pathway

The quinate utilisation pathway converts quinic acid (QA) to protocatechuic acid (PCA). It is under carbon catabolite repression with glucose being the preferred carbon source. The highly inducible quinate pathway shares intermediates with the low level constitutive shikimate pathway (Figure 1.1). The shikimate pathway converts phosphoenol pyruvic acid to chorismic acid the precursor of aromatic amino acids and other aromatic compounds. Maintenance is therefore essential for growth and despite competition for intermediates with the induced quinate pathway. Additionally, organisation of the regulation of the quinate pathway ensures that these common intermediates, present under all conditions, do not cause induction of the quinate pathway in the absence of quinate. The relationship between the two pathways and the genes involved has proved to be the key to understanding the regulation of the *qut* system. The pioneering work on the two pathways was carried out in *Neurospora crassa* by Prof. N. H. Giles and colleagues and a brief review of the results and ideas derived from this work is given

1.3.1 Quinic acid utilisation (qa) and Shikimate pathways in *N.crassa*

Three enzymes catalyse the conversion of quinic acid (QA) to protocatechuate (PCA). The presence of quinic acid (QA), dehydroquinate (DHQ) or dehydroshikimate (DHS) induces these enzymes. Induction is subject to carbon catabolite repression (Chaleff, 1974a: Chaleff, 1974b). The three enzymes encoded by the genes *qa3* (quinate dehydrogenase), *qa2* (3-dehydroquinase) and *qa4* (dehydroshikimate dehydratase) and regulated by the *qa-1* locus. All mapped to a tightly linked gene cluster on chromosome VII (Case & Giles, 1975; 1976). Northern blot analysis showed that the regulatory locus *qa-1* consists of two genes designated *qa-1F,* encoding an activator, and *qa-1S,* encoding a repressor (Huiet, 1984). Two additional genes induced by quinic acid, identified by northern blot analysis and designated *qa-x* and *qa-y* (Patel *et al.,* 1981). Their function was unknown. A pentafunctional polypeptide carries out steps 1-5 in the shikimate pathway. The protein encoded by the *aromA* gene is active as a dimer (Giles *et al.,* 1967; Jacobson *et al.,* 1972; Lumsden & Coggins, 1977; 1978). Mutations in the biosynthetic 3-dehydroquinase (bDHQase) function of the *AROM* protein constitutively expressed the QA pathway enzymes, due to internal accumulation of DHQ. This observation led to the hypothesis that the pentafunctional *AROM* dimer had a channelling function to separate the internal pools of DHQ and DHS produced by the shikimate pathway. The channelling function may prevent the induction of the QA pathway (Giles, 1978). However, this interpretation was challenged by Duncan *et al.,* (1987) who held that the genes encoding the five catalytic functions fused into a single gene to enable strict co-ordinate gene expression.

Figure 1.1 **Quinic acid utilisation and shikimate pathways in *Aspergillus nidulans***



The figure shows the quinic acid and shikimate pathways in *Aspergillus nidulans*.

The shared intermediates indicated by the verticle arrows crossing the dotted line Enzymes responsible for the converion of intermediates are shown in italics.

Genes encoding the quinate enzymes are also indicated. Numbers *1-5* represent steps in the shikimate pathway carried out by the AROM pentafunctional enzyme.

1.3.2 The *AROM* protein of the shikimate pathway in *A.nidulans*

The quinate and shikimate pathways in *A.nidulans* as in *N.crassa* share two intermediates DHQ and DHS. DHQ, DHS or QA induces expression of the *qut* pathway genes (Hawkins *et al.,* 1982; Kinghorn & Hawkins, 1982). The *aromA* locus in *A.nidulans* has been cloned and sequenced. The pentafunctional *aromA* gene appears to have evolved by multiple gene fusion of the monofunctional genes. Separate enzymes are found in many prokaryotes (Charles *et al.,* 1985; Charles *et al.,* 1986; Hawkins, 1987; Hawkins & Smith, 1991). Since separate genes in prokaryotes encoded the same five functions as *aromA* gene, this suggests that co-ordinate expression of the enzymes is not the essential feature that determined their fusion. The original experiments (Gaertner *et al.,* 1970) utilised AROM protein that was partially degraded (Duncan *et al.*, 1987) and lacked zinc ions necessary for full activity (Lambert *et al.,* 1985) casting further doubt upon the channelling hypothesis. Resolving the distribution of intermediates between the quinate and shikimate pathways clearly required more work. The proposed channelling function of the AROM dimer was examined *in vivo*. *A.nidulans* was transformed with an expression vector containing the wild-type *A.nidulans* *aromA* gene. The host *qutE*$^{-}$; *qutR*^{c} strain constitutively expresses the *qut* genes but is unable to utilise quinic acid as a carbon source due to the loss of *qutE* encoded catabolic dehydroquinase activity. Transformants were selected for growth on quinic acid. Strains that over-expressed the *aromA* gene grow, although poorly, on quinic acid as a sole carbon source. This can only occur if DHS leaks from the AROM dimer. In further experiments an *aromA*$^{-}$ host strain, that requires aromatic amino acid supplemented media for growth, was transformed with the wild-type *aromA* gene. Transformed strains growing on unsupplemented media with glucose as a carbon source were recovered. The selected strains when transferred for growth on quinic acid as a carbon source exhibited reduced levels of the quinate pathway enzymes. This supports the proposal that the over-expressed AROM protein sequesters the shared intermediates DHQ and DHS, that also serve as the signals for *qut* pathway induction. This indicates that the dimer can utilise intermediates formed by the quinate pathway (Lamb *et al.,* 1991). These experiments infer that the shikimate pathway proceeds despite competition from the induced *qut* enzymes due to the localised concentration of active sites in the pentafunctional AROM dimer that ensures the supply of intermediates. The internal concentration of DHQ and DHS normally appear too low to bring about induction of the *qut* pathway. However, low levels of DHQ and DHS that leak from the AROM dimer may be responsible for the basal level expression of the *qut* gene cluster.

The shikimate pathway as a target for pathway engineering: The shikimate pathway is present in all prokaryotes, microbial eukaryotes and plants studied. The pathway is absent animal higher eukaryotes and is thus a potential target for specific chemotherapy upon microbial infection.

The end point of the shikimate pathway is chorismic acid, the precursor for aromatic amino acids. It also forms a branch point for an extensive range of metabolites including metal chelators, vitamins E and K, folic acid, ubiquinone and plastiquinone. The pre-chorismic acid shikimate pathway also branches at 3-deoxy-d-arabino-heptulosonate-7-phosphate and 3-dehydroquinate giving rise to 3-hydroxy 5-aminobenzoate and 3-aminobenzoate respectively. These are precursors for anamycins a varied and complex group of compounds with a wide range of biological properties, including antibiotic, anti-viral and anti-tumour activities (Review Bentley, 1990). The ability to flux intermediates from the highly inducible quinate pathway into the naturally low level shikimate pathway make this system, with its myriad of branching points, an attractive target for pathway engineering of medical and biotechnological importance.

1.3.3 **Quinic acid utilisation (*qut*) in *A.nidulans***

A.nidulans utilises quinic acid as a sole carbon source for growth. Induction of the quinic acid pathway is subject to carbon catabolite repression. The genes responsible for quinate utilisation and regulation of the pathway map to a tightly linked cluster on chromosome VIII (Hawkins *et al.*, 1982: Kinghorn & Hawkins, 1982: Hawkins *et al.*, 1984). The genes encoding the three catabolic enzymes *qutB* (quinate dehydrogenase), *qutE* (catabolic dehydroquinase) and *qutC* (dehydroshikimate dehydratase) were isolated using heterologous probes from *N.crassa* (Hawkins *et al.*, 1985).

The two genes *qutA* and *qutD* were originally identified as potential regulatory genes. Mutants of *qutD* were non-inducible for all three enzyme activities, interpreted as loss of function of a transcriptional activator (Hawkins *et al.*, 1984). However, the phenotype of *qutD* mutant strains is reversible by growth at low pH, where non-ionised quinic acid enters the cell by diffusion. The *qutD* gene (homologous to *N.crassa qa-y*) encodes a quinic acid permease. Recovered amongst revertants of mutant *qutD* strains, selected for growth on quinate at normal pH 6.5, were weakly constitutive strains. These proved to be double mutant strains. The second locus was designated *qutR*. A single recessive mutations at the *qutR* locus produces high levels of the quinate pathway enzymes in the absence of quinate, suggesting that *qutR* encodes a repressor (Whittington *et al.*, 1987). Rare dominant non-inducible *qutR* mutant strains were also isolated. These alleles may encode a repressor protein that retains its repression activity but has lost its ability to respond to the presence of an inducer. The dominant non-inducible *qutR* mutant strains are unstable (4.5 revertants in 10^5 colonies plated), reverting readily, with the majority of revertants being recessive constitutive mutants (Grant *et al.*, 1988).

All mutant *qutA* strains fail to induce the quinate pathway in the presence of quinate. The wild-type *qutA* gene from *A.nidulans* has been cloned and sequenced. The gene encodes an 824 amino

acid protein that showed some identity to the *GAL4* transcriptional activator. The homology to the *GAL4* gene product and the phenotype of the *qutA* mutant strains strongly suggests that the gene encodes a transcriptional activator for the quinic acid pathway (Beri *et al.*, 1987).

A molecular model for *qut* pathway regulation: The model described below uses the basic principles of the Douglas & Hawthorne model, (1966) for galactose utilisation regulation in *S.cerevisiae*. The model incorporates details of the molecular action of the regulatory proteins determined from molecular studies (Ma & Ptashne, 1987a: Ptashne, 1988), and the genetic and biochemical study of *qut* regulation.

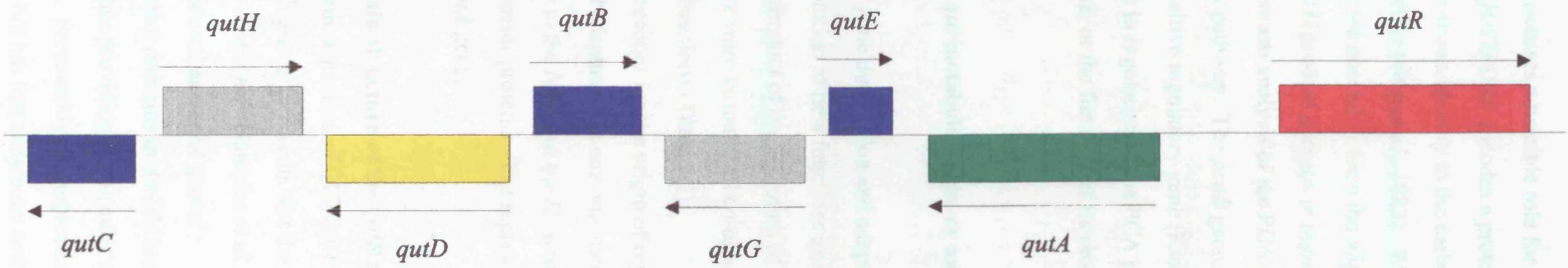
The *qutA* activator protein binds to a promoter element in each of the quinate inducible genes. In non-inducing conditions the *qutR* repressor protein interacts with the activator preventing its transcriptional function. In the presence of quinate, DHQ or DHS the repressor binds an inducing molecule causing it to release the activator to promote transcription. Quinate utilisation leads to a drop in concentration and the repressor and inducer dissociate. This results in the repressor binding to the activator protein.

All *qutA*$^-$ mutant strains isolated were non-inducible in the presence of quinate. However, the mutants fall in one of three classes when combined in heterozygous diploid strains with the wild-type *qutA*$^+$ strain. Heterozygous diploid strains with recessive alleles grow on quinic acid at a rate indistinguishable from a homozygous *qutA*$^+$ diploid strain. The mutant allele presumably produces no functional protein. The single wild-type allele produces sufficient protein for normal regulation. Heterozygous diploid strains with semi-dominant alleles show a definite but reduced ability to grow on quinic acid as a sole carbon source. This presumably reflects the formation of a partially functional mutant protein that interferes with some aspect of *qut* pathway regulation by the wild-type activator. Rare dominant *qutA*$^-$ alleles in the heterozygous diploid do not grow on quinic acid as sole carbon source. This infers a mutant protein that totally blocks induction of the *qut* pathway by the wild-type activator protein (Grant *et al.*, 1988).

The *qut* cluster in *A.nidulans*: The genetic map of the *qut* gene cluster has the gene order *qutC, D, B, E, A, R* (Grant *et al.*, 1988). Cloning and sequencing the entire 19kb *qut* gene cluster identified each of the known genes and confirmed the order determined by the genetic map. However, the sequence analysis also identified two new ORFs designated *QUTG* and *QUTH*, not identified by mutation (Hawkins *et al.*, 1988). Northern blot analysis showed that both ORFs produced quinate inducible messages. Figure 1.2 shows the order and organisation of the genes in the *A.nidulans qut* gene cluster.

The *QUTG* ORF (homologous to the *N.crassa qa-x* gene) produces a protein that shows identity to bovine myo inositol monophosphatase. A phosphatase enzyme does not form part of the quinate

Figure 1.2 **The quinic acid utilisation (*qut*) gene cluster of *Aspergillus nidulans***



The figure shows the oganisation of the quinic acid (*qut*) gene cluster on chromosome VIII in *A.nidulans*

Arrows indicate the direction of transcription. *qutC, B & E* encode the *qut* enzymes.

qutD encodes the permease. *qutG & H* encode quinate inducible genes, function unknown.

qutA the activator & *qutR* the repressor.

pathway, but Section 1.4 discusses a possible role for a phosphatase in *qut* pathway regulation (Lamb *et al.*, 1990). The *QUTH* ORF encodes a protein containing a putative Zinc cluster DNA binding motif. The protein shows identity to the carboxyl terminus of the DNA directed DNA polymerase of hepatitis B virus (Ono *et al.*, 1983). Whilst the similarity is low, it may be significant as no manual inserted gaps are needed to form the alignment. Together, these observations suggest regulatory role for the *QUTH* product perhaps in regulation of the PCA pathway genes (Lamb *et al.*, 1992). Genetic and biochemical analysis of the PCA pathway shows that its regulation is independently from the *qut* pathway. The *pcaB* gene, that maps to chromosome VIII proximal to the *qut* cluster, is a putative positive regulatory gene (Kuswandi & Roberts, 1992). Therefore, if the *QUTH* product plays a role in regulation of the PCA pathway it may be an additional factor for the expression of the *pcaB* gene, or the function of its product.

1.4 The Evolution of the *qut* metabolic pathway and its regulatory genes

Enzyme recruitment by gene duplication and adaptation has long been recognised as being an essential feature in the evolution of new functions and metabolic pathways (Jensen, 1976). The *qut* pathway appears to be the product of the evolution of a limited number of functional units. Components of the pathway were virtually all co-opted from related pathways or functions that evolved pathway specific functions (Table 1.1).

The same process may account for the origin of regulatory genes, with regulatory genes being evolved by duplication and adaptation of enzyme encoding genes. The *AROM* pentafunctional polypeptide shows identity to the *N.crassa qa-1S* (Anton *et al.*, 1987) and *A.nidulans qutR* (Hawkins *et al.*, 1992) repressor proteins. The repressor appears to have evolved from the 3' region of the multifunctional *aromA* gene.

1.4.1 Evolution and domain structure of the *qutR* repressor protein

Comparison of the protein sequences of QUTR (*A.nidulans*) QA1-F (*N.crassa*) and *AROM* from *A.nidulans, N.crassa* and *S.cerevisiae* reveals that the repressor shows some identity with the three carboxyl terminal domains of *AROM* (Hawkins *et al.*, 1992). Figure 1.3 shows the proposed evolutionary origin of the *aromA* and *qutR* genes.

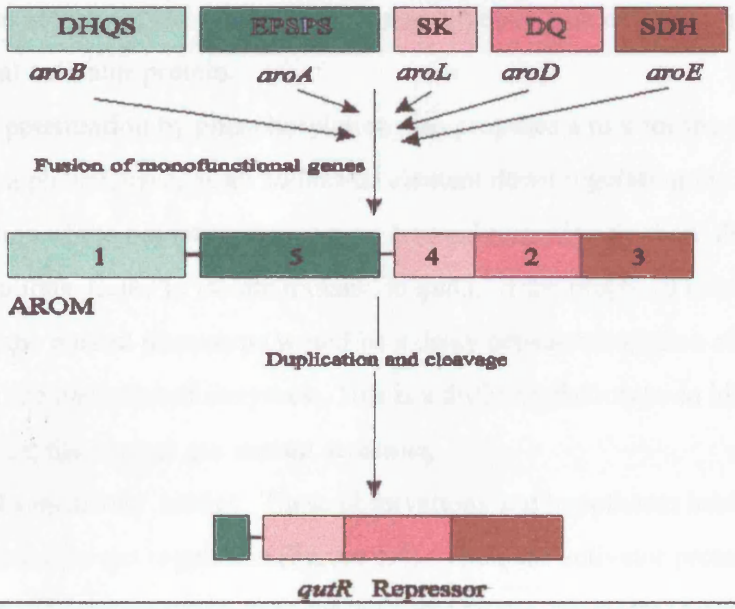
The three carboxyl terminal domains in *AROM* have enzymatic activity and are able to recognise and bind DHQ and DHS, thus providing a molecular mechanism for the recognition of the inducing molecules by the repressor. Presumably, the repressor protein has retained the ability to recognise and bind the intermediates but has lost enzymatic activity towards them. Indeed, the *qutR* protein

Table 1.1 **Evolution of the *qut* metabolic pathway**

Gene	Product	Protein identity
qutB	Quinate dehydrogenase	Bacterial & eukaryotic shikimate dehydrogenases, including AROM (Hawkins *et al.*, 1988)
qutE	Catabolic dehydroquinase	Bacterial type II dehydroquinase (Garbe *et al.*, 1991: Kleanthous *et al.*, 1992)
qutD	Quinate permease	A modified sugar transporter (Geever *et al.*, 1989)
qutG	Function unknown	A modified sugar phosphatase (Lamb *et al.*, 1990: Neuwald *et al.*, 1991)
qutH	Function unknown	Hepatitus B virus DNA polymerase (Lamb *et al.*, 1992)
qutC	Dehydroshikimate dehydratase	No significant homology in the database

The table shows the products of the *qut* gene cluster and their identity with other gene products. The quinate pathway is thought to have evolved by adoption of genes encoding similar functions that have evolved pathway specific activity.

Figure1.3 **Evolutionary origins of the *aromA* and *qutR* genes**



The figure shows the prposed evolutionary origins of the *aromA* gene by multiple gene fusion. The figure also shows the subsequent duplication and cleavage of *aromA* that led to the evolution of the *qutR* reporessor (Hawkins *et al.*, 1992). DHQS = dehydroquinate synthase (step 1). EPSPS = enol pyruvate shikimate phosphate synthase (step 5). SK = shikimate kinase (step 4). DQ = dehydroquinase (step 2). SDH = shikimate dehydrogenase (step 3). The step numbers on the figure refer to the steps in the pathway shown in Figure 1.1.

contains the bDHQase domain catalytic site with a single amino acid (aa) substitution of an active lysine residue for an inactive arginine (Chaudhari *et al.*, 1991). It is probable that QA, DHQ and DHS exist in equilibrium in the cell at low concentrations and either DHQ or DHS may be the actual inducer.

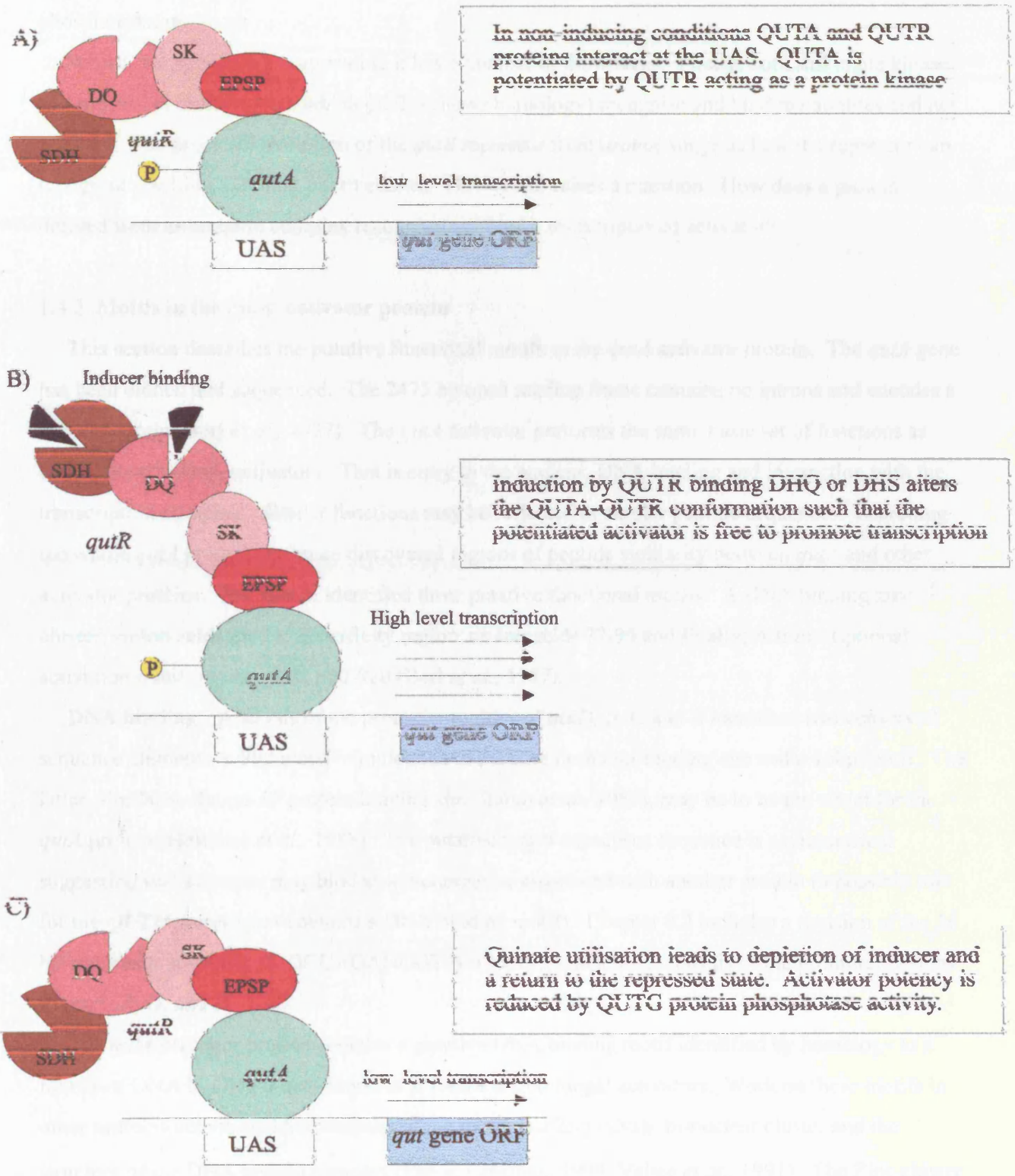
The homology suggests another role for the repressor, in addition to explaining how the repressor binds the inducing metabolites at altered catalytic sites, derived from the *AROM* pentafunctional enzyme. The shikimate kinase domain represented in the repressor protein includes a motif that potentially binds purine nucleotides. This suggests the repressor may have an *in vivo* protein kinase activity with the *qutA* activator being the substrate - an appealing hypothesis when considering the following points.

Phosphorylation plays a key role in transcription regulation. Phosphorylation of the *S.cerevisiae ADR1* transcriptional activator prevents expression of the *ADH2* gene (alcohol dehydrogenase), but does not prevent activator-DNA binding (Cherry *et al.*, 1989: Taylor & Young, 1990). In contrast to the ADR1 protein phosphorylated forms of the GAL4 activator are related to activation of transcription (Mylin *et al.*, 1990: Sadowski *et al.*, 1991). The transition between initiation and elongation transcription complexes involves phosphorylation of the RNA polymerase II terminal subunit (Payne *et al.*, 1989; Laybourn & Dahmous, 1990; O'Brien *et al.*, 1994). The proposed protein kinase activity of the *qutR* repressor may potentiate an otherwise ineffective *qutA* transcriptional activator protein.

Activator potentiation by phosphorylation also proposes a role for the *QUTG* product, with its homology to a phosphatase, as an additional element down regulating the *qut* gene expression by de-phosphoralation of the potentiated activator. Normal screening method, that is looking for QA non-utilising mutations, failed to isolate mutants in *qutG*. If the proposed role for the *QUTG* product is correct, then the mutant phenotype would be a delay between depletion of the substrate and repression of the *qut* induced enzymes. This is a difficult phenotype to identify and would not be recovered using the normal *qut* mutant screening.

A revised regulatory model: These observations and hypotheses lead to further revision of the molecular model for *qut* regulation (Figure 1.4). The *qutA* activator protein binds to the promoter of the *qut* genes. The repressor binds to the activator on the promoter (supporting evidence from *N.crassa*, Baum *et al.*, 1987). In non-inducing conditions the repressor binding prevents transcriptional activation; at this point the repressor phosphorylates the *qutA* protein potentiating its activity. In the presence of quinic acid basal level expression of the *qut* pathway enzymes ensures the rapid accumulation of the inducing metabolites, which the repressor binds. This may lead to dissociation of the activator and repressor or alter conformation of the binding, such that the activator interacts with the transcription complex to promote gene expression. As the metabolite

Figure 1.4 **Molecular model for the regulation of the *qut* gene cluster**



The figures show a proposed molecular model for the regulation of *qut* pathway by QUTA and QUTR. The putative role of QUTG as an additional down regulator is shown. The homologous domains of the repressor to AROM are labelled. (P) represents the proposed phosphate groups added to the activator.

pool becomes depleted the repressor releases the inducing molecules reinstating repression. *QUTG* protein phosphatase action may facilitate repression by reducing the potency of the activator by de-phosphorylation.

Whilst this hypothesis is appealing it has a number of drawbacks. Firstly, both shikimate kinase and inositol phosphatase (to which *QUTG* shows homology) recognise and bind metabolites and not proteins. The proposed evolution of the *qutR* repressor from *aromA* suggests how the repressor can recognise and bind inducing intermediates. However it raises a question. How does a protein derived from an enzyme complex recognise and bind a transcriptional activator?

1.4.2 **Motifs in the *qutA* activator protein**

This section describes the putative functional motifs in the *qutA* activator protein. The *qutA* gene has been cloned and sequenced. The 2475 bp open reading frame contains no introns and encodes a 825 aa protein (Beri *et al.*, 1987). The *qutA* activator performs the same basic set of functions as other transcriptional activators. That is entry to the nucleus, DNA binding and interaction with the transcription complex. Similar functions may be reflected in similar peptide sequences. Searching the whole *qutA* protein sequence discovered regions of peptide similarity between *qutA* and other activator proteins. The search identified three putative functional motifs: A DNA binding zinc cluster, amino acids 49-76; Specificity region, amino acids 77-90 and finally; A transcriptional activation motif, amino acids 800-820 (Beri *et al.*, 1987).

DNA binding: Analysis of the promoter regions of *qutD, B, G* and *E* identified two conserved sequence elements, a 9bp motif homologous to the core promoter binding site and a 16bp motif. The latter, similar to the *qa-1F* protein binding site (Baum *et al.*, 1987), may be to be the target for the *qutA* protein (Hawkins *et al.*, 1988). The putative target consensus sequence is asymmetrical suggesting that activator may bind as a monomer or associated with another protein (a possible role for the *QUTH* protein that contains a DNA binding motif). Chapter 8.2 includes a revision of the 16 bp consensus sequence (5' GCCAGANCGTTCTNCC 3') including analysis of the promoter regions of *qutA, R, G*, and *H.*

The *qutA* activator protein contains a putative DNA binding motif identified by homology to a family of DNA binding motifs found in at least a dozen fungal activators. Work on these motifs in other proteins determined the structure of the motif as a Zn_2 $(Cys)_6$ bi-nuclear cluster and the structure of the DNA protein complex (Pan & Coleman, 1990: Vallee *et al.*, 1991). The Zinc cluster motif binds DNA. A region immediately adjacent, the pathway specific motif, was proposed to provide the specificity of binding to the *qut* genes.

Transcriptional activation motifs: Classification of transcriptional activation motifs in transcription factors falls into three groups: acidic (GAL4/GCN4), proline rich (CTF/NF1) and

glutamine rich (SP-1) (Reviewed by Mitchel & Tjian, 1989). Lower eukaryotic regulatory transcription factors most commonly encode the acidic class motif. Analysis of the *qutA* protein sequence suggested that an acidic class motif capable of forming an ampiphatic helix lies at the carboxyl terminus.

Despite these regions of peptide similarity, transcriptional activators show little overall amino acid sequence identity. In addition to their common functions activators encode motifs unique to the their particular role, such as inducer, or repressor, binding sites. Comparison of the entire activator protein sequence to catalytic proteins with related functions may provide new insight into the evolution and function of transcriptional regulators.

Application of this logic led to the discovery of the origins of the *qutR* repressor from the *aromA* gene. If the *qutR* repressor protein binds the activator protein, a search for a protein that AROM binds may prove valuable in identification of the evolutionary origins of the *qutA* activator.

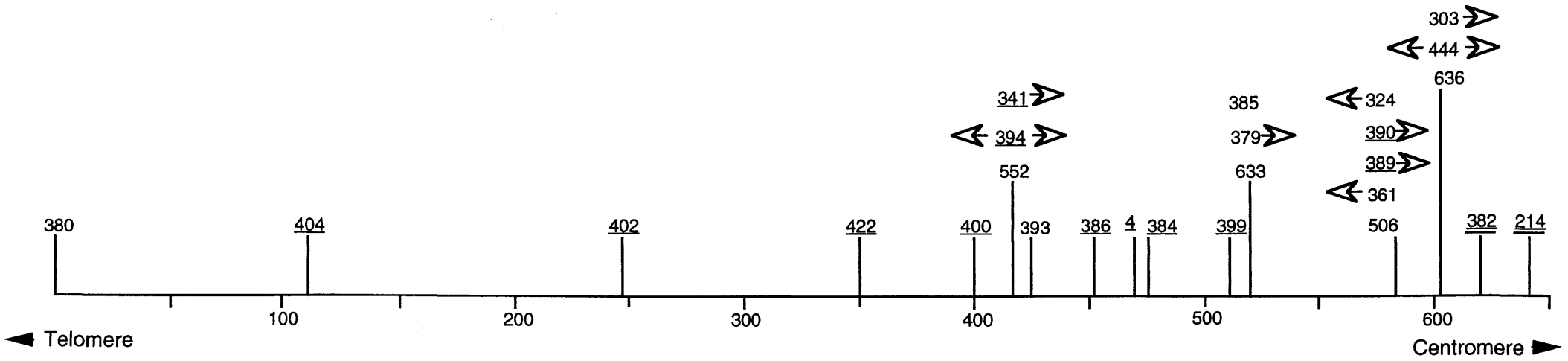
Aims of the work

The central aim of the work is to discover functional motifs in the *qutA* activator protein. This was to be achieved by DNA sequence analysis of a series of *qutA* mutant alleles and interpretation of the inferred protein changes. Certain *qutA* mutations generate a dominant phenotype in the heterozygous diploid. This was interpreted as being due to mutant protein blocking the action of wild-type activator by binding to the DNA target sequence and acting as a repressor. Analysis of these alleles should provide data about DNA binding and transcriptional activation motifs within the activator protein.

Figure 1.5 shows the genetic map of the *qutA* locus constructed using the data shown in Appendix 2. The initial aim of the work was to secure an alignment of the genetic and physical maps. Mutant alleles for DNA sequence analysis of a particular region of the *qutA* gene could then be selected using the data from the genetic map. This avoids the need to sequence the whole of the *qutA* gene for each mutant allele.

Analysis of randomly generated non-inducible *qutA* mutations provides the opportunity for discovery of actual functional regions of the protein containing no recognised motif. Additionally, the randomly produced mutations may provide insights into the domain structure and function of the *qutA* activator.

Figure 1.5 **The genetic map of the *qutA* locus**



The figure shows the genetic map of the *qutA* locus and the sites of mutations calculated from the genetic data given in Appendix 2. The scale on the lower line is wild-type *qutA* recombinants per million ascospores plated. Recessive alleles are denoted by plain text. Semi-dominant alleles are denoted by underlined text. Dominant alleles are underlined twice. Mutations surrounded by arrows represent those that are separated by less then ten recombinants per million and therefore their location cannot be accurately determined. The orientation of genetic map is from Grant *et al.*, (1988).

Chapter Two

Materials and Methods

2.1 Strains

2.1.1 *Aspergillus nidulans* nomenclature

All strains used are derivatives of the Glasgow collection (Pontecorvo *et al.*, 1953). The nomenclature used is based on those published by Clutterbuck, (1973). The strain numbers '*R21*' refer to the stock collection held in the Department of Genetics at the University of Leicester (This collection was transferred to the care of Dr. A. R. Hawkins, Department of Biochemistry and Genetics, University of Newcastle Upon Tyne in September 1994).

Genetic loci and mutants are designated by italicised one to four letter symbols for example *y* = yellow conidia and *qut* = quinic acid utilisation. Non-allelic loci with the same primary symbols are distinguished by the following capital letter for example *qutA* and *qutR* represent different genes affecting quinic acid utilisation. Alleles of the same gene are distinguished by a number following the locus letters for example *qutA4* or *qut A214*. Wild-type alleles are distinguished using a superscript plus symbol for example $qutA^{+}$. Where mutant alleles of a gene are referred to in general they are denoted with a superscript minus symbol that is $qutA^{-}$. Superscript letters are also used to describe specific properties for example $qutR^{c}$ refers to mutants of the *qutR* gene which constitutively express the quinic acid utilisation genes. The genotype of the strains described in this thesis is given in Table 2.1.

Mutant allele symbols

Conidiospore colour:	*yA* = yellow conidiospores	*wA* = white conidiospores
	The wild-type (yA^{+}; wA^{+}) has green conidia.	
Nutritional requirements:	*paba* = Para-amino benzoic acid	*pyro* = Pyridoxine
Carbon source utilisation	*qut* = quinic acid utilisation	

Table 2.1 **Strain list**

Strain		Genotype	Strain		Genotype
R21	*1*	*yA2; pabaA1*	***R153***	*1*	*wA3; pyroA4*
qutA4	*2*	*wA3; pyroA4; qutA4*	***qutA214***	*3*	*wA3; pyroA4; qutA214*
R21qutR16^{c}	*4*	*y A2;paba A1; qut R16^{c}*	***qutA303***	*4*	*wA3; pyroA4; qutA303*
qutA324	*4*	*wA3; pyroA4; qutA324*	***qutA341***	*4*	*wA3; pyroA4; qutA341*
qutA361	*4*	*wA3; pyroA4; qutA361*	***qutA382***	*4*	*wA3; pyroA4; qutA382*
qutA444	*4*	*wA3; pyroA4; qutA444*	***qutA379***	*5*	*wA3; pyroA4; qutA379*
qutA384	*5*	*wA3; pyroA4; qutA384*	***qutA385***	*5*	*wA3; pyroA4; qutA385*
qutA386	*5*	*wA3; pyroA4; qutA386*	***qutA389***	*5*	*wA3; pyroA4; qutA389*
qutA390	*5*	*wA3; pyroA4; qutA390*	***qutA393***	*5*	*wA3; pyroA4; qutA393*
qutA394	*5*	*wA3; pyroA4; qutA394*	***qutA399***	*5*	*wA3; pyroA4; qutA399*
qutA400	*5*	*wA3; pyroA4; qutA400*	***qutA402***	*5*	*wA3; pyroA4; qutA402*
qutA404	*5*	*wA3; pyroA4; qutA404*	***qutA422***	*5*	*wA3; pyroA4; qutA422*
qutA423	*5*	*wA3; pyroA4; qutA423*			
qutA506	*5*	*yA2;pabaA1;qutR16^{c}; qutA506*			
qutA552	*5*	*yA2;pabaA1;qutR16^{c}; qutA552*			
qutA633	*5*	*yA2;pabaA1;qutR16^{c}; qutA633*			
qutA636	*5*	*yA2;pabaA1;qutR16^{c}; qutA636*			

In the Table the non-italicised letters (*1-5*) designate the reference for, or source of the strains listed.

1 = Fantes & Roberts, 1973. *2* = Hawkins *et al.*, 1982 *3* = Hawkins *et al.*, 1985

4 = Grant *et al.*, 1988 *5* = C.F. Roberts

2.1.2 *Esherichia coli* strains and cloning vectors

Strain	Genotype	Source
JM103	*supE; thi1; endA1; hsdR4; sbcB15; strAÆ (lac -proAB); F' (traD36; proAB$^+$; lacI^q; lacZÆM15).*	A.R. Hawkins [Messing *et al.*, (1981)]
DH5-αsupE44;	*ÆlacU169; hsdR17; recA1; gyrA96; thi1; relA1; ø80lacZÆM15).*	B. M. Wilkins [Gibco BRL Life Technologies]

Allele symbols

proAB	= Proline requirement,	*endA*	= DNA specific endonuclease I
lacI	= Repressor of *lac* operon,	*hsdR*	= Host specific endonuclease R
lacZ	= β-D-galactosidase,	*gyrA*	= Gyrase Nalidixic acid resistance
recA	= General recombination,	*relA*	= Regulation of RNA synthase
strA	= Streptomycin resistance,	*supE*	= Suppresser of amber (UAG) mutations
thi	= Thiamine requirement,	*sbcB*	= Exonuclease I suppresser of *recB*, *recC*

Phage	Genotype	Source
M13mp18	*lac* Z'; M13mp18 polylinker	A.R. Hawkins, [Yanisch-Perron, (1985)]

Plasmid	Genotype	Source
pUC18	*lac* Z'; M13mp18 polylinker	B. M. Wilkins, [Yanisch-Perron, (1985)]

2.2 *A.nidulans* Growth Media

Distilled water: All solutions were made in autoclaved deionised water unless otherwise specified.
Sterilisation of media: All media was sterilised by autoclaving at a pressure of 15 pounds per square inch for 20min, unless otherwise specified.

Minimal media (MM)

Defined minimal medium was based on that described by Pontecorvo *et al.*, (1953), as modified by Roberts, (1963). A 1 in 10 dilution of the stock salts, containing trace elements. (10x Stock salts contained: 6g $NaNO_3$; 0.52g KCl; 1.52g KH_2PO_4; 1ml Trace elements solution; made up to 1ltr with water. The pH of the stock salts was adjusted to pH6.5 using NaOH pellets and dilute sodium hydroxide solution (Trace elements solution contained per litre: 1g $FeSO_4.7H_2O$; 8.8g

$ZnSO_4.7H_2O$; 0.4g $CuSO_4.5H_2O$; 0.15g $MnSO_4.4H_2O$; 0.1g $Na_2B_4O_7.10H_2O$; 0.05g $(NH_4)_6Mo_7O_{24}$; made up to 1ltr with water).

Liquid culture medium contained the wetting agent Tween 80, diluted 10^5 (v/v) to prevent mycelium from clumping and clinging to the sides of the flasks. Minimal medium was solidified with 1.8% (w/v) agar to produce minimal agar plates (MA). After autoclaving the media was allowed to cool to 50-60°C, when sterile $MgSO_4$ (52g/200ml filter sterilised) was added to a final concentration of 10mM. Appropriate vitamin supplements were added and a carbon source provided as described.

Malt Extract Agar (MEA)

This complex medium was used as a complete medium to produce vigorous growth and abundant formation of vegetative conidiospores was prepared with the following quantities: 20g Malt extract (Oxoid); 20g Glucose (Oxoid); 18g Agar; 1g Bactopeptone per litre of media. Appropriate nutritional supplements were added after autoclaving.

Supplementation of media

i) Vitamins	Stock solution	Final concentration
para-amino benzoic acid (PABA)	0.1g/100ml	1µg/ml
pyridoxin (PYRO)	0.05g/l00ml	1µg/ml
ii) Carbon source	Stock solution	Final concentration
Glucose	1M	20mM
Glycerol	1M	50mM
Quinic Acid pH6.5	20% (w/v)	0.5% (w/v)

Carbon source stock solutions in distilled water were sterilised by autoclaving at a pressure of 15 pounds per square inch for 15min.

2.3 *Escherichia coli* growth media

Bacterial growth media were similar to those described in Miller, (1972) and were used as liquid broth or solid media when containing agar to a final concentration of 2% (w/v).

Baltimore Biological Laboratories (BBL) media.

Components per litre of media: 5g Trypticase (BBL), 5g NaCl, 2.5g $MgCl_2.7H_2O$, 18g (6g for overlays) Agar.

Luria Broth (LB) Media

Components per litre of media: 10g Tryptone (Oxoid), 5g Yeast extract (Oxoid), 5g NaCl, Adjusted to pH7.0 with NaOH.

Yeast extract/tryptone (YT) media

Components per litre of media: 16g Yeast extract (BBL), 10g Tryptone (Oxoid), 5g NaCl, the media was adjusted to pH 7.2 with 5M NaOH.

Antibiotic selective media

	Stock solution	Final concentration
Ampicillin	20mg/ml	100mg/ml
Nalidixic acid	40mg/ml (50% ethanol)	25mg/ml

Stock solutions were stored at 4°C for no more than 1 month.

β-Galactosidase selection of recombinant plasmids and phage

	Stock solution	Final concentration
IPTG	25mg/ml	50mg/ml
X-gal	20mg/ml	40mg/ml

2.4 Methods for handling *Aspergillus nidulans*

Growth and storage of strains: All cultures were grown at 37°C. Stock cultures of strains were maintained on MEA slants. Long term storage of strains was on silica gel (Roberts, 1969).

2.4.1 Preparation of suspensions of conidiospores

A conidial suspension in 1ml Tween/saline solution was made by taking up conidia in loop of a sterile solution, until the suspension became cloudy. A sample of 100-200µl of the suspension was spread over the surface of supplemented agar plates and incubated at 37°C for 2-3 days until vigorous conidiation had occurred. Plates were stored at 4°C until required.

Plates were flooded with 10ml sterile Tween/saline and the conidia gently removed from the surface of the plate by scraping the mycelium with a sterile glass rod. The resultant conidial suspension was transferred to a 30ml Sterilin tube and centrifuged at 3000rpm in a Baird & Tatlock

type IV bench top centrifuge for 5min to pellet the conidia. The conidia were washed in sterile phosphate buffer (20mM KH_2PO_4; 50mM Na_2HPO_4; 50mM NaCl; 0.4mM $MgSO_4$; pH7.2) and stored in this solution for up to 4 weeks at 4°C. The concentration of the conidial suspension was determined by spectrophotometry and comparison to a standard curve. The suspension generally held in the range of 10^7-10^8 conidiospores per ml.

2.4.2 **Sexual crosses**

Meiotic recombination was used in the genetic mapping of *qutA*$^-$ mutant strains. Sexual crosses were set up between haploid strains with complementary nutritional requirements and whenever possible contrasting conidiospore colour markers, that is the majority of crosses involved strains derived from *R153* (*wA3*; *pyroA4*), which produces white conidiospores and *R21* (*yA2*; *pabaA1*), which produces yellow conidiospores.

Conidiospore suspensions from the two strains to be crossed were and 1ml of each suspension was mixed together on a vortex mixer. The spores were then sedimented by centrifugation described in Section 2.4.1 the liquid phase decanted and the conidial pellet resuspended in approximately 500ml phosphate buffer and spread as a single streak on a thick MA (40ml) plate supplemented with glucose. The plates were then part sealed with masking tape, to prevent drying and incubated at 37°C for three to four weeks. The mature cross was viewed under a dissecting microscope and several cleistothecia, seen as black shiny spheres, were picked with a sterile mounted needle. The cleistothecia were rolled gently across a 2% (w/v) water agar plate to remove the coating of buff coloured Hülle cells and contaminating hyphae and conidiospores. Individual cleistothecium were then crushed on the side of a 1.5ml Eppendorf tube containing 1ml of phosphate buffer (pH 7.2) to release the haploid ascospores. Each fruiting body contains 10^4-10^5 asci each of which in turn contains 8 haploid progeny. Therefore each suspension contained 10^5-10^6 ascospores. The ascospores were used in the identification and isolation of recombinant strains. Growth tests were performed to determine the phenotype of recombinants.

A.nidulans is homothallic, therefore a proportion of the cleistothecia are the product of self-fertilisation. It was noted that hybrid cleistothecium tended to be larger. To distinguish selfed and hybrid fruiting bodies a sample of spore suspension from a number of different fruiting bodies were separately streaked in a radial pattern were across MEA plates and incubated for 3 days at 37°C. Selfed fruiting bodies gave rise to single colour streaks whilst hybrid fruiting bodies produced three colours of colony: white ($y^-w^-$and y^+ w^-), yellow (y^-w^+) and the wild-type green recombinant (y^+w^+) in the ratio 2:1:1.

2.4.3 **Construction of heterokaryons and diploid strains**

This method was used to determine the phenotype of *qutA*$^-$ mutants with both wild-type and other *qut* mutant strains (Roper, 1952).

Conidiospores from two auxotrophic strains with different nutritional requirements were mixed and plated on a MA plate supplemented with glucose and incubated for 18hr.

Weak mycelial growth occurred and blocks of agar were cut from the growing mycelial edge. A dozen blocks were transferred to a fresh MA glucose supplemented plate with the mycelial growth in contact with the agar, the plates were incubated for 3-5 days. In some cases there was no growth and the process was repeated. In other cases mycelial growth was observed around the transferred section blocks of agar were again cut from the edge of the strongly growing heterokaryon and laid mycelia side down, four per plate on fresh MA glucose supplemented plates and incubated for 3 days.

Conidia from heterokaryon growth were scraped off the plate using a spatula and suspended in 10ml sterile water. Aliquots of the suspension were added to 25ml volumes of glucose supplemented MA and then poured into petri dishes, allowed to cool and incubated for 2-3 days. Conidia are separated by suspension in the distilled water and molten media, diploid conidia grow through the media before any heterkaryon growth can occur. The volume of conidial suspension plated is altered to allow separate diploid colonies to form. Diploid strains were selected and used in growth tests.

2.4.4 **A colourimetric 'spot test' for the production of PCA**

The PCA 'spot test' is a used for the detection of protocatacheuic acid (PCA). This method was described by Partridge *et al.*, (1972) and was developed by C.F. Roberts. The principle of the test is that non-induced mycelium fail to carry out the this overall reaction, since the enzymes of the qut pathway are absent or at low levels. However mycelium induced by growth on QA, or if the enzymes of the pathway are constitutively expressed produce PCA. The procedure has been proved to be a sensitive and reliable test to detect QA enzyme activity in mycelium grown on solid media by the detection of PCA (Grant *et al.*, 1988).

A heavy innoculum of conidiospores was patched onto the surface of supplemented MA plates and incubated for approximately 18hr at 37^{o}C. Control strains wild-type *R153* and constitutive *R21:qutR16*^{c} strains were plated on QA supplemented MA plates as a positive control and upon glycerol supplemented plates as a non-inducing carbon source when testing for constitutivity.

Samples of the mycelium were taken from agar plates using an inverted yellow Gilson pipette tip as a bore to remove a consistent sized block of agar. The mycelial plugs were then removed from

the tip using a sterile toothpick and each placed in a separate small (5ml) glass test tubes. To each tube 1ml of the following freshly prepared reaction mixture was added; 5ml (DMSO); 5ml (0.4M) potassium glycinate pH9.2; 5ml (20% w/v) QA pH6.5; 1.3ml (1%) $FeCl_3$; 8ml H_2O.

The samples were then incubated at 30°C overnight. The mycelium is permeabalised by the DMSO. PCA produced by the activity of the QA pathway enzymes is detected as a purple colouration of the liquid and agar plug. The colour develops due to complex formation between the PCA and Fe^{3+} ions. With a negative result the solution and agar remain a straw yellow colour.

2.4.5 **Genetic mapping of mutant *qutA* alleles**

The tightly linked (2% Recombination frequency, RF) recessive constitutive mutation *qutR16* was used as a flanking marker to order mutant sites in the *qutA* gene by meiotic analysis.
The genetic mapping exercise required *qutA* mutants in the constitutive strain to serve as markers in *qutA* locus. New *qutR^c qutA* double mutants were isolated in *R21qutR16*. UV mutagenesis and filtration enrichment (Armitt *et al.*, 1976) was carried out to recover new *qut* mutants by their failure to grow on QA and the new *qutA* mutants identified by their failure to complement a well defined *qutA* mutation (*R153, qutA303*). The procedure isolated four new *qutA* mutations for the genetic mapping exercise *qutA506, 552, 633* and *636*. The alleles were all recessive when combined with a *qutA*$^+$ wild-type allele in diploid strains.

Sexual crosses were set up between each of the four "new" *qutA*$^-$ strains and the other *qutA* mutants in the *R153* background as described in Section 2.4.2. Hybrid cleistothecia were individually identified and the contents from 5-10 such fruiting bodies pooled as a suspension in each inter allelic cross:-

Viable count: The pooled ascospore suspension was diluted 10µl:90µl in phosphate buffer and equal portions plated on 4 or more MEA plates supplemented with PABA, PYRO and glucose. The plates were incubated for 2-3 days at 37°C then the number of colonies were counted. The viable count gave the number of ascospores in the suspension.

Selective plating for wild -type *qutA* recombinants: Aliquots of the remaining ascospore suspension were plated in 10 or more MA plates supplemented with PABA, PYRO and QA, incubated as before and the number of colonies recorded. Selection for growth on QA recovers *qutA*$^+$ strains resulting from recombination between the different *qutA*$^-$ alleles in the cross. The RF was determined by calculating the number of QA$^+$ recombinant colonies recovered divided by the total number of ascospores plated. The reversion rates of mutants mapped was less than 1 in 10^6.

The PCA "spot test" in ordering mutant *qutA* alleles: A sample randomly selected QA^+ colonies were patched on MA plates supplemented with PABA, PYRO and glycerol and incubated at 37°C for 18hr. The PCA spot test was performed to determine the number of colonies that constitutively expressed the QA utilisation enzymes; those that contained the $qutR16^c$ allele and those that contained the $qutR^+$ allele. Recombination yielding $qutA^+$strains is represented in Figure 2.1.

In the majority of cases the alleles could be reliably located to the left or to the right of one another with respect to the *qutR16* marker. In a some crosses so few $qutA^+$ recombinants were recovered that alleles could not be ordered or separated. The resulting genetic map is presented in Figure 1.5.

2.4.6 Preparation of *Aspergillus nidulans* genomic DNA

A.nidulans genomic DNA was isolated by two methods. The first, a preparation from mycelia, yielded better quality and a greater quantity of DNA than the second "mini-prep" method from conidiospores. However the mini-prep method was used more extensively as it was quicker and the quantity and quality of the DNA yielded by this method proved to be sufficient for use as template in PCR amplification.

DNA preparation from mycelia: A 500ml baffled conical flask containing 200ml liquid MM was inoculated with 2 x 10^8 conidiospores prepared as in Section 2.4.1. and incubated in a gyratory incubator shaker at 37°C for 20hr. Mycelia were then collected by filtration onto 3MM Whatmann paper washed with 1x stock salts and dried by blotting with more Whatmann paper. The mycelia were wrapped in aluminium foil and stored at -20°C.

To disrupt the cell wall, a 1-2g sample of frozen mycelia were ground to a fine powder under liquid nitrogen using pre-cooled mortar and pestle. The powder was suspended in 10ml of extraction buffer (100mM Sodium acetate, 2.5mM EDTA, 4% SDS) by gently shaking in a polypropylene tube.

Proteins were removed by phenol extraction in phenol resistant polyallomer tubes. The samples were centrifuged in a Beckman JA20 rotor at 10,000rpm for 10min. The cellular debris formed a layer at the interface of the upper aqueous phase and the lower organic phase.

The upper aqueous phase was removed carefully, transferring it to a fresh tube. The volume was made up to 14ml with extraction buffer. The tubes were then centrifuged in a Beckman SW 40 Ti swing out rotor at 20,000rpm for 1hr. The polysaccharides form a hard colourless pellet at the base of the tube. The supernatant is again removed to a fresh Sterilin tube.

Figure 2.1 **Genetic mapping of *qutA* alleles**

The figure shows the rational by which *qutA* mutations were mapped relative to each other and the *qutR* gene.

A) Recombination events that for wild-type *qutA* would be expected to have an excess of wild-type *qutR* over constitutive *qutR*. The majority are inducible.

B) Recombination events that for wild-type *qutA* would be expected to have an excess of constitutive *qutR* over wild-type *qutR*. The majority are constitutive.

The DNA was then purified by a caesium chloride gradient centrifugation. The separation of high m.w. DNA from low m.w. RNA and other cellular contaminants depends upon the buoyant density of DNA in the CsCl gradient. This was done by adding 0.4g of caesium chloride/ml CsCl to 9ml of sample (2.4M). This was then layered onto 4ml of 5.8M CsCl, 0.1M EDTA, in a 14ml polyallomer tube. The sample was centrifuged at 20,000rpm for 18hr in a SW 80 Ti rotor, forming a CsCl gradient in the tube. The tube contains a cloudy upper layer of cellular debris which was removed and discarded. The clear layer below contains the high molecular weight DNA. The majority of the RNA forms a pellet at the base of the tube. The clear layer was removed to a glass universal.

The CsCl was removed by dialysis against Tris-EDTA (10mM Tris-HCl pH7.5; 1mM EDTA pH7.5) buffer . The solution was mixed with 5ml Tris-EDTA buffer and transferred to a dialysis membrane bag, sealed and dialysed for 3hr against 3x 1.5l aliquots of dialysis buffer. The DNA was recovered by ethanol precipitation and then resuspended in 200µl sterile distilled water.

The DNA sample was digested with excess DNase free RNase, 1mg of RNase added to 100ml of sample and incubated at 37°C for 3 hr. This method yielded approximately 500-1000mg DNA per sample and was typically resuspended in 500µl of sterile distilled water and stored for up to 6 months at -20°C.

DNA preparation from conidiospores: The following method was found most useful for the rapid preparation of genomic DNA as template for the polymerase chain reaction (PCR) and subsequent analysis of *qutA*$^-$ strains. The procedure was described by Beri *et al.*, (1987)

All the following steps were carried on ice. A suspension of was centrifuged, the supernatant decanted and the conidial pellet resuspended in 0.8ml cold spermidine lysis buffer (SLB 5mM Spermidine; 100mM KCl; 10mM EDTA; 10mM Tris-HCl pH7.5; 250mM Sucrose). The suspension was transferred to an Eppendorf tube and the conidia pelleted, the removed and the pellet resuspended in a further 0.8ml SLB. The slurry of spores was transferred to a 3ml glass universal capped bottle containing 3g 0.45mm glass beads and violently agitated in a vortex mixer to disrupt the spores. Breakage was monitored under the light microscope by the proportion of empty "ghost" spores seen. A further 0.8ml SLB was added and the contents allowed to settle

The liquid phase was then transferred to two Eppendorf tubes, 0.1 volume of 10% (w/v) SDS added and the mixture incubated at 60°C for 20min after mixing. From this stage onwards samples were mixed in Eppendorf tubes. A 0.25 volume of 4M sodium acetate (pH6.0) was added, mixed gently by inversion of the tube to avoid breakage of DNA molecules by sheering which was then held on ice for 30min. Debris was pelleted by centrifugation for 5min and the supernatant removed to fresh tubes.

Proteins were removed by digestion with Proteinase K added to a final concentration of 200mg/ml, and RNA digested with DNAase free RNase at a final concentration of 100mg/ml. Incubation was at 37°C for 1hr. The DNA was precipitated by the addition of 0.1 volume of 4M sodium acetate (pH6.0) and 0.6 volume of isopropanol, the solution was mixed gently by inversion and then the DNA pelleted by centrifugation for 10min. The supernatant was removed, the pellet washed in 70% (v/v) ethanol, which was then removed and the pellet first dried under vacuum then resuspended in 200ml sterile distilled water. Samples were gently mixed as a series of phenol extractions were performed until no debris was seen at the interface. The DNA was re-isolated as described above and the pellet resuspended in 200ml sterile water. The mini-prep method yielded 150-200mg DNA per sample. An aliquot of the DNA sample was diluted to 1ng/ml for use as template in the PCR.

2.5 Methods for handling *Escherichia coli*

E.coli strains were used in the preparation of plasmid and phage DNA for the cloning and sequencing of PCR amplified DNA.

2.5.1 Growth and storage of strains

All cultures were grown at 37°C. For short-term storage the organism was streaked on Luria Agar plates. A single colony was used to re-streak fresh solid media at three month intervals. Long-term storage of strains was carried out as described in Maniatis *et al.*, (1982). A single colony was used to inoculate 5ml of LB and grown to stationary phase by incubation with shaking at 37°C overnight. The cells were pelleted by centrifugation at 5,000rpm for 5min in a Baird & Tatlock mark IV bench top centrifuge at 4°C and the pellets resuspended in 5ml LB. The cells were pelleted as before, resuspended in 1.5ml LB and 0.4ml aliquots mixed with 2.5ml sterile 50% (v/v) glycerol in 2ml screw cap tubes. The mixture was rapidly frozen at -80°C in a dry ice/ethanol bath and stored at -80°C. The cells were revived by transferring a small sample from the surface of the frozen material to a Luria agar plate. The stored sample was immediately returned to -80°C.

2.5.2 DNA mediated transformation of *E.coli*

The *E coli* strains JM103 were transformed with phage DNA and DH5α when using the plasmid vector. The method is based upon the method described by Mandel and Higa, (1970) and adapted as Maniatis *et al.*,(1982).

A 1.5ml overnight stationary phase culture of *E. coli* in LB was used to inoculate 30ml of LB pre-warmed to 37°C, which was vigorously aerated in a gyratory incubator at 37°C until the culture

reached an OD of 0.45-0.6 at a wavelength of 600nm. Approximately 10ml of the culture was chilled on ice for 15min and the cells were pelleted at 2,500rpm for 10min in a Baird & Tatlock mark IV bench top centrifuge at 4°C. At this stage cells may be used to prepare frozen stocks as described below. Cells for immediate use in transformation were resuspended in 10ml of ice-cold 100mM $CaCl_2$ and incubated on ice for 10min. The cells were pelleted again before resuspension in 1ml ice-cold 0.1M $CaCl_2$. The cells were incubated on ice for 60min at which time they were competent for transformation. Frozen stocks of competent cells were prepared by centrifugation and then the pellet resuspended in 5ml 50mM $CaCl_2$, 5% (v/v) Glycerol. The mixture was dispensed in 500µl aliquots in Eppendorf tubes and "snap frozen" by transfer to a dry ice/ethanol bath. The cells were then stored at -80°C for up to one month.

A sample of 0.5-1µg recombinant plasmid of phage DNA for transformation was mixed with 200µl of competent cells (either freshly prepared or thawed slowly from a frozen stock) and incubated on ice, with occasional agitation for 50min. The mixture was heat-shocked by incubation at 42°C for 2min and returned to ice for 20min. A volume of 10ml of LB pre-warmed at 37°C was added to each sample of cells and the mixture incubated at 37°C for 90min. Suitable samples (100µl) of the transformation mixture were spread onto LA plates containing the appropriate antibiotic for selection of transformants and incubated overnight at 37°C. Positive controls were provided by transforming *E.coli* with the original. Samples of cells not exposed to transforming DNA were plated to provide a negative control.

2.5.3 **Preparation of plasmid DNA from *E.coli***

The small-scale preparation of bacterial plasmid DNA was carried out according to the procedure of Birnboim & Doly, (1979). This method was used to prepare vector plasmid pUC18 for use in cloning PCR amplified DNA and recombinant plasmid DNA for use template in sequencing reactions.

Approximately 1.5ml of an overnight culture of *E.coli* was transferred to an Eppendorf tube and the cells harvested by centrifugation for 2min in a MSE Microcentaur centrifuge. The supernatant was removed by aspiration and the pellet resuspended in 100µl of an ice cold TEG solution (25mM Tris.HCl, pH8.0, 10mM EDTA, 50mM glucose). The sample was held at room temperature for 5min and 200µl of a freshly prepared solution of 0.2M NaOH, 0.1% (w/v) SDS was added. The contents were gently mixed inversion and incubated on ice for 5min. A 150µl aliquot of an ice-cold potassium acetate solution (60ml of 5M potassium acetate, 11.5ml of glacial acetic acid, 28.5ml of Q-water, pH4.8) was added to the tube, mixed by inversion and incubated on ice for 5min. The tube was centrifuged for 5min to yield a clear supernatant which was transferred to a fresh tube.

Protein was removed by a series of phenol extractions, the tube was mixed then centrifuged for 2min and the aqueous phase transferred to a fresh tube. RNA was removed from the sample by the addition of 1µl DNAase free RNase A (10mg/ml) and incubation at 37°C for 1hr. This was followed by a further phenol extraction. The DNA was recovered by ethanol precipitation. The samples were mixed and incubated at room temperature for 15min. The nucleic acids were resuspended in 50µl sterile distilled water and stored at -20°C.

2.5.4 **Transfection of *E coli* with phage M13mp18**

Approximately 1mg of recombinant phage DNA was mixed with 200µl competent cells of *E. coli* strain JM103. The mixture was incubated on ice for 30min and then heat shocked at 42°C for 2min. The cells were then mixed with 3.5ml BBL soft agar at 45°C, 200µl of a logarithmic phase culture of *JM103*, 20µl of 20mg/ml IPTG and 20µl of 20mg/ml X-gal. The mixture was poured onto a BBL plate and allowed to solidify. The plates were incubated overnight at 37°C to allow for growth of phage and plaques, the required recombinant phage plaques are colourless, non-recombinant phage form blue coloured plaques.

Analysis of phage M13 recombinants: Phage from colourless plaques were picked with a Pasteur pipette, mixed vigorously with 2ml 2x YT broth containing a 50-fold dilution of a logarithmic phase culture of *E. coli* JM103. The culture was vigorously aerated in a gyratory incubator shaker at 37°C for 6hr. The cells were centrifuged at room temperature for 5min, the supernatant was removed and stored. An aliquot of the pelleted cells was resuspended in 5% (v/v) glycerol and stored for later use. Double stranded replicative form phage DNA was isolated from the remainder of the pelleted cells. The resulting DNA samples were separated by electrophoresis in a 0.7% agarose gel at 90V for 3hr and the gel stained with 5µg/ml ethidium bromide. A negative control phage DNA sample prepared using blue, wild-type, plaques was also loaded on the gel. Recombinant phage yield DNA migrating at a slower rate than the original phage due to the presence of an inserted sequence.

Preparation of recombinant single stranded phage M13 DNA as a template for sequencing reactions: The supernatant from the phage infected culture described above contains single stranded phage particles released from infected cells of *E.coli*. Phage particles were precipitated by addition of 250µl 30% w/v PEG 8000, 1.6M NaCl. The suspension was vigorously mixed, incubated for 15min at room temperature and then centrifuged for 5 min. The supernatant was removed from the pellet using a Gilson pipette and the rest of the liquid removed by a number of cycles of centrifugation and pipetting. When the pellet was completely dry, it was resuspended in 200µl sterile distilled water. To remove the phage protein coat a phenol extraction was performed. The aqueous layer was transferred to a clean tube containing 20µl 3M sodium acetate pH 5.5 and

400µl 100% ethanol. After mixing the solution was placed at -80°C for 15min and centrifuged for 5min the pelleted DNA was recovered as described above and stored -20°C until required for sequencing.

2.6 General molecular techniques

The majority of molecular procedures were carried out using 1.5ml Eppendorf tubes. Where a procedure required centrifugation this was carried out in a Centaur microcentrifuge unless otherwise stated.

Digestion of DNA with restriction endonuclease: Digestion with type II restriction endonucleases was performed with reaction buffers provided by the manufacturer and at the temperature and time recommended. The reaction was typically carried out in a volume of 10-50µl and involved the digestion of less than 1µg of DNA. Wherever possible the reaction was stopped by heat inactivation of the restriction endonuclease, otherwise the reaction volume was made up to at least 100µl and phenol extractions performed with the DNA subsequently being recovered by ethanol precipitation.

Removal of 5' phosphate groups by phosphatase treatment: The frequency of the re-ligation of vector DNA molecules with cohesive ends can be reduced by treatment with alkaline phosphatase to remove the 5' terminal phosphate group thus preventing self-ligation. Following digestion with the selected restriction endonuclease 10 units of Bovine alkaline phosphatase was added and incubated at 37°C for 1hr. Circularisation of the vector can then only occur by insertion of the un-treated DNA fragments which provide the 5' terminal phosphate at each end for the ligation reaction to occur.

Addition of 5' phosphate groups by T4 DNA Kinase: The primers used in PCR lack the 5' phosphate group necessary for ligation reactions. This phosphate group is provided by treatment with T4 Kinase. The prepared amplified fragment is incubated with kinase buffer (66mM Tris-HCl pH7.6; 1mM ATP; 1mM spermidine; 10mM $MgCl_2$; 15mM DTT; 0.2mg/ml BSA) and 2 units of T4 DNA kinase at 37°C for 1hr. The enzyme can be heat denatured by incubation at 65°C for 15 min.

Fill-in of recessed ends of DNA fragments: It has been observed Taq polymerase used in PCR often adds an additional non-templated nucleotide to the end of a synthesised DNA strand. This means that the majority of PCR amplified DNA fragments have a single base sticky end. This single

base overhang must be "filled in" before the fragment can be cloned into a vector digested with a restriction endonuclease which leaves blunt ends (for example *Sma*I). Filling in of 3' recessed ends of DNA fragments generated by restriction endonuclease digestion was carried out by incubating the recovered fragment in 16µl sterile water, 2µl 10x "fill in" buffer (10x "Fill in" Buffer: 500mM Tris.HCl pH 7.2, 100mM $MgSO_4$, 10mM DTT, 500µg/ml BSA, stored in 100 ml aliquots at -20°C), 2µl Chase mix (Chase mix: 10mM Tris.HCl pH 8.0; 1mM EDTA; 25 mM Each dNTP; Chase mix was stored in 100 ml aliquots at -20°C) and 1 unit of the Klenow fragment of DNA polymerase I at 37°C for 30min. The reaction as stopped by phenol extraction and the DNA recovered by ethanol precipitation and resuspended in the required volume of water.

Ligation of DNA molecules: DNA ligase was used to covalently join either "sticky" or "blunt" ended DNA fragments. Ligations were carried out in reaction volumes of 10-20ml containing 1x ligation buffer (10x ligation buffer pH7.0: 50mM Tris-HCl, 10mM $MgCl_2$, 10mM DTT, 10mM ATP). The reactions were incubated at 16°C for 12-20hr and recombinant molecules recovered by transformation into *E.coli.*

The molar ratio of the different DNA species in a ligation reactions is important and condition which maximise the yield of recombinant molecules can be calculated from theoretical analysis (Dugaiczyk *et al.*, 1975). In practice this means a 10x molar excess of insert to vector was used in the reactions.

Protein removal by phenol extraction: Phenol equilibrated with 0.5M Tris-HCl, pH7.4, and 1%(v/v) with respect to 8-hydroxy quinoline was mixed with an equal volume of chloroform/isoamyl alcohol (24:1). Equal volumes of phenol were added to the aqueous DNA solution and mixed as directed in the protocols. The tube was then centrifuged for 2min. The upper aqueous layer was then removed to a fresh tube, taking care not to disturb the precipitated protein at the interface. The extraction was repeated until no further debris was observed. Traces of phenol in the aqueous phase were then removed by the same method by extraction with chloroform alone.

Ethanol precipitation of DNA: Ethanol precipitation was used to recover DNA from aqueous solutions. Generally this involved the addition of 0.05 volume of 4M sodium acetate and 2.2 volumes of 100% ethanol the solution was then mixed gently to precipitate the DNA. Variation were introduced in the salts used and their relative concentrations, the volume of ethanol or isopropanol and the time and temperature allowed for precipitation depending upon the particular size DNA fragments recovered. The DNA was then pelleted by centrifugation and the pellet washed

in 70% (v/v) ethanol centrifuged, the ethanol removed and the pellet dried under vacuum. The dried pellet was redissolved in sterile distilled water.

2.6.1 **Agarose gel electrophoresis**

The size and concentration of the horizontal slab gels was chosen with regard to the nature of the samples analysed. Routinely 0.8% agarose gels were used with low melting point gels being used when the required band was to be cut from the gel and DNA recovered. In general mini-gels of 30ml volume were used for rapid electrophoresis, larger gels (200ml) were used for a more accurate resolution. The agarose was dissolved in 1x TAE was melted and poured into the appropriate gel bed and allowed to set. The comb providing the sample wells was removed and the gel submerged in 1x TAE and electrophoresis performed at 15-100V for 1-20hr depending on the time available resolution required. High resolution was achieved at lower voltages whilst fast separation was achieved with the higher voltages.

Tris acetate Electrophoresis Buffer (50x TAE: 2M Tris; 1M CH_3COONa (sodium acetate); 50mM EDTA; the solution was adjusted to pH8.2 with glacial acetic acid).

Recovery of DNA fragments from agarose gels: This procedure was used to separate PCR products from the oligonucleotide primers and unincorporated nucleotides, as well as for the isolation of digested vector molecule fragments for use in the cloning of PCR products.

A number of different methods of extracting DNA from agarose gel were attempted the most reliable of which are detailed below.

i) 'Death wish': Two gels were prepared the first was loaded with the DNA sample and electrophoresis carried out. The second gel had a slot cut in it slightly larger than the size of the wells of the first gel. The bottom and front of the slot is lined with a single layer of dialysis membrane (previously boiled in 10mM Tris-HCl pH7.5; 1mM EDTA). When the first gel shows sufficient separation of the DNA sample, this can be visualised using ethidium bromide and a hand held short wavelength UV transilluminator (Exposure to UV is kept to a minimum reducing UV damage to DNA), the required band is cut from the gel and placed in the slot in the second gel. Electrophoresis of the second gel is now carried out in a dark room. The DNA is impacted on the surface of the membrane by the electrical current. When the dialysis membrane is loaded with DNA, this can be seen using the hand held UV lamp, the membrane is removed (with the current still running) using a pair of plastic handled (Millipore) tweezers and quickly transferred to an Eppendorf tube. The edge of the membrane is trapped between the lid and side of the Eppendorf and centrifuged for a few seconds. The DNA, no longer held to the membrane by the electrical potential, is now in solution at the bottom of the tube. This liquid is removed and an ethanol

precipitation performed to recover the DNA. The membrane may be viewed under UV illumination to check if all the DNA has been removed, if not then sterile water is added and the tube shaken then centrifuged as before. This procedure typically yielded 100-200ng purified DNA from a gel loaded with 1mg of the fragment required.

ii) 'Gene Clean': This method was originally described by Vogelstein & Gillespie (1979). The procedure and reagents are supplied commercially by BIO101 as their 'Gene Clean' kit. Samples were loaded onto a 0.8%-1.0% (w/v) low gelling temperature agarose gel (FMC Sea Plaque agarose) and electrophoresis carried out to give suitable separation of the fragments involved. The piece of agarose containing the required fragment was removed using a scalpel, transferred to an Eppendorf tube and compacted by centrifugation to determine the approximate volume of the gel slice. A 2-3x volume of saturated NaI (supplied with the 'Gene Clean' kit) was added to the tube. The agarose was melted by incubation at 55°C for 8min, when 5µl of an aqueous solution of glass powder (Glassmilk, as supplied with the 'Gene Clean' kit) was added and incubated for 5min on ice to allow binding of DNA to the glass surface. The Glassmilk was pelleted by a 5sec centrifugation, the supernatant discarded and the pellet washed four times in 900ml of NEW WASH buffer. The washed pellet was then resuspended in 25µl sterile water and incubated at 55°C for 3min to elute the DNA from the Glassmilk. The glassmilk was sedimented by centrifugation for 1-2min spin at high speed and the supernatant containing the DNA transferred to a fresh tube, this procedure was repeated with a further 10µl of sterile water, giving a final volume of 35ml. An aliquot of this solution was run on a gel with a known amounts of marker DNA to estimate the recovery of fragment DNA. This method typically yielded 500-750ng DNA from a 1mg sample. The DNA recovered was of suitable quality for use as template in sequencing reactions.

2.7 Polymerase Chain Reaction (PCR)

PCR was used to amplify specific regions of the *qutA* gene from genomic DNA of wild-type and mutant strains. The PCR products were subsequently used in SSCP analysis (Section 2.8), providing template for cloning and direct sequencing of PCR amplified regions of the *qutA* gene.

PCR primers were designed using the "Primers" program (Lowe *et al.,* 1990) and the sequence of the wild-type *qutA*$^+$ published by Beri *et al.,* (1987). Due to the difficulty in cloning PCR products, primers were also designed by inspection of the Beri sequence to incorporate restriction sites within the oligonucleotides. The oligonucleotides were synthesised by the Department of Biochemistry at Leicester University and John Keyte Department of Biochemistry Nottingham University.

Preparation of primers: The primers were delivered in a 2ml, 35% ammonia solution. Only one quarter of this solution was prepared at any one time as the ammonium solution stabilises the primers. A 0.1 volume of 2M sodium acetate and 3.0 volumes 95% ethanol was added to 0.5ml of the primer solution. The contents was vigorously mixed and placed at -80°C for 15min. The tube was centrifuged for 20min to pellet the DNA. The pellet was washed in 70% ethanol centrifuged and the ethanol removed. The pellet was dissolved in 125µl sterile water. The concentration of the primer was then estimated by spectrophotometry at 260nm and an aliquot diluted to produce the 10mM solution used in PCR. The same solution of primers were also diluted further to 0.5pM to be used for direct sequencing of PCR products.

The PCR method: The PCR method used was derived from Saiki *et al.*, (1988). Reactions were carried out in a total volume of 50ml in 0.5ml Eppendorf tube. To ensure maximal yield of product a series of experiments were carried out to determine the optimum quantity of genomic DNA template (ranging from 1pg to 1 µg) and the volume of primer (1-15µl). The reaction conditions for each set of primers were also optimised individually with respect to annealing temperatures, extension times and cycle number. The PCR reaction were carried out in a Techne PHC-3 thermal cycler.

The reaction mixture contained: 1ng/ml *A.nidulans* genomic DNA (10µl), 7.4x PCR-mix (6.75µl), Each Primer (5µl), Taq polymerase 0.5 units, made up to 50µl with sterile deionised water mixed and the sample overlaid with mineral oil to prevent evaporation. In the preparation of template for SSCP analysis (Section 2.8) the amplified DNA was labelled by the addition of 0.5µl of ^{32}P-dCTP.

PCR-mix contains: 167µl 2M Tris-HCl pH8.0; 83µl 1M $(NH4)_2SO_4$; 33.5µl 1M $MgCl_2$; 3.4µl 10mM EDTA pH8.0; 7.5µl 100mM Each dNTP; 85µl 10mg/ml BSA; 3.6µl β-mercaptoethanol; 270µ H_2O. The mixture was vigorously mixed then separated into 35µl aliquots and stored at -20°C.

Reaction conditions

Primer set 1: Cycles 1-5; Denaturing (94°C for 5min), Annealing (55°C for 1.5min), DNA synthesis (72°C for 1min). Cycles 5-25; Denaturing (94°C for 1.5min), Annealing (55°C for 1.5min) DNA synthesis (72°C for 1min).

Primer set 2: Cycles 1-5: Denaturing (94°C for 5min), Annealing 58°C for 1min), DNA synthesis 72°C for 1.5min). Cycles 5-25; Denaturing (94°C for 1.5min), Annealing (58°C for 1min), DNA synthesis (72°C for 1.5min).

Primer set 3 Cycles 1-5; Denaturing (94°C for 5min), Annealing (56°C for 1.5min), DNA synthesis (72°C for 1.5min). Cycles 5-25; Denaturing (94°C for 1.5min), Annealing (56°C for 1.5min), DNA synthesis (72°C for 1.5min)

A longer denaturing period was used in the first five cycles to ensure complete separation of genomic DNA strands. All the PCR reactions were completed with a final DNA synthesis step of 5min. DNA synthesis step to ensure all PCR product was full length. The PCR product was then removed from under the mineral oil using a Gilson pipette and transferred to a fresh tube.

Preparation of PCR product for sequencing: Five separate amplification reactions were carried. The products were then combined and separated from excess primer and unincorporated nucleotides by electrophoresis in a low melting point agarose gel. The resultant band was cut from the gel and the DNA purified by the Gene Clean method (Section 2.6.1). This procedure typically yielded around 500-750ng template DNA. It was found that 200ng of an 800bp fragment was a sufficient quantity of template for a sequencing reaction.

2.8 Single Stranded Conformational Polymorphism (SSCP)

The preliminary alignment of the genetic map of the *qutA*$^-$ alleles to the physical map of the gene was achieved by SSCP analysis. This procedure was carried out using PCR products of genomic DNA from a number of different *qutA*$^-$ mutants.

The procedure was that described previously by Orita *et al.*, (1989) and was developed by Mackay *et al.*, (1993). The method allows the detection of single base changes in DNA, due to the altered mobility of single stranded DNA fragments. The conformation of the DNA strands is dependent upon the sequence, alterations of which are sometimes reflected in an alteration of conformation and so mobility on a non-denaturing polyacrylamide gel.

Restriction endonucleases identified by analysis of the published QUTA sequence (Beri *et al.*, 1987) were used to digest 8µl of the ^{32}P dCTP labelled PCR product. The restriction endonucleases were chosen to give a range of fragment sizes between 30-500bp (the sizes most likely to show detectable alterations in mobility). Once the digestion was completed 10µl of denaturing solution was added (270µl 10% SDS; 40µl 500mM EDTA; 940µl H_20; fresh denaturing solution was prepared for each set of reactions). The digestion products can then be stored at -20°C until ready to load on the non-denaturing gel. When ready to load the gel, 3µl of the solution was mixed with an equal volume of formamide dye (95% (v/v) Formamide; 20mM EDTA pH7.5; 0.05% (w/v) Xylene cyanol; 0.05% (w/v) Bromophenol blue; a 40 ml stock solution was prepared and stored at -20°C). A 5µl aliquot of the sample was loaded directly as double stranded DNA. Meanwhile a second portion was mixed with the formamide dye boiled for 5min to denature the DNA and 5µl loaded onto the gel in a single stranded form.

The preparation of the plates to contain the gel for SSCP was the same as that described for sequencing gels (Section 2.9.4). The gel was run at 4°C and a constant power of 30W for 5-6hr. The gel plates were then separated the gel transferred to 3MM Whatmann paper and dried at 80°C under vacuum. The dried gel was put in a cassette with x-ray film and stored at -80°C overnight (or longer) before developing to reveal areas exposed to ^{32}P.

Non-denaturing polyacrylamide gel: 9ml Acugel (38% acrylamide, 2% N', N'-bisacrylamide); 6ml 10x TBE buffer (10x Tris Borate Electrophoresis (TBE) Buffer; 87.2g Tris; 44g Boric acid; 7.44g EDTA; made up to 1ltr with deionised water); 12ml 50% (v/v) Glycerol; 33ml Distilled water. The gel was then degassed under vacuum before the following were added. 1ml 10% ammonium persulphate; TEMED 21µl. The gel was poured and allowed to set for a minimum of 2hr.

2.9 DNA sequencing

The template for sequencing reactions came from a number of different sources. The first method of sequencing attempted involved production of single stranded template by PCR. The failure of this method led to PCR product being cloned into the two sequencing vectors phage M13mp18, plasmid pUC18. The use of pUC18 as a cloning vector was preferred to M13mp18 as the plasmid preparation handling and stability were more secure. Cloning the PCR products was eventually superseded as a route towards DNA sequencing, when the direct preparation of PCR product of a quantity and quality suitable for sequencing was achieved.

2.9.1 **Direct sequencing for single stranded PCR products**

This method was adapted from Gyllensten & Erlich (1988) and Bachmann *et al.*, (1990) and the standard 'sequenase' protocol. It allows the direct sequencing of PCR amplified DNA, incorporating a step for generation of ssDNA by PCR. One of the primers used in the original PCR is used to generate the single stranded template (half PCR), the other for the sequencing reactions on this template.

i) Preparation of the dsDNA template

The DNA fragment to be sequenced was amplified by PCR as dsDNA using the standard PCR protocol (Section 2.7). The PCR product was then purified as described in section 2.6.1.

ii) Preparation of the ssDNA template

The dsDNA fragment produced in step i) was used as template for the production of ssDNA, in a PCR reaction containing only one primer (half PCR). Half PCR reaction contained: 200-500ng dsDNA; 1x PCR mix, 1mM Primer, 1 unit Taq polymerase; made up to a final volume of 50 µl with sterile distilled H_2O. The DNA was amplified for 16 cycles under the same conditions as the last 16 cycles of a normal PCR. The products were removed from under the mineral oil and a selective precipitation was performed to recover the ssDNA. A 0.6 volume of 5M ammonium acetate and 1.4x volume of isopropanol was added, mixed and then held at -80°C for 15min. The tube was centrifuged for 20min, the pellet was washed with 70% ethanol dried and resuspended in 7µl water. This selective precipitation is designed to precipitate the single stranded PCR product but not any excess primer or unincorporated nucleotides which may affect the sequencing reaction. An agarose gel was loaded with 1µl of the sample to determine the amount of product.

iii) Sequencing reactions

The remaining 6µl DNA was incubated with 1µl primer "B", 2µl 5x sequenase buffer (200mM Tris-HCl pH7.5; 100mM $MgCl_2$; 250mM NaCl), 1µl 10% nonidet P-40. The mixture was boiled for 3min and snap frozen in a dry ice/IMS bath. Radiolabelling was carried out as the mixture thawed at room temperature, by the addition of : (1µl) 0.1M Dithiothreotol; (2µl) labelling mix diluted 1:5 (7.5mM dCTP, dGTP, dTTP); (0.5µl) $\alpha^{35}S$ dATP; (2µl) T_7 DNA polymerase (diluted 1:8 =1 unit per reaction). The reaction mixture was divided equally between the four termination mixes(80mM Each dNTP; 50mM NaCl; 8mM of the appropriate ddNTP). After incubation at 37°C for 5min, 2µl chase mix (0.25mM Each dNTP; 50mM NaCl; 50mM Nonidet P-40) was added and a further 5min incubation carried out at 37°C. 4µl sequencing dye was added

and the reactions left on ice or stored at -20°C until ready to load onto the Urea/polyacrylamide gel. The samples were boiled for 5min immediately prior to loading.

2.9.2 **Sequencing of cloned PCR products**

DNA amplified by PCR was cloned the sequencing vectors described below and template for sequencing was prepared as described in for ssDNA M13mp18 clones and for dsDNA pUC18 clones. The procedure and regents used for sequencing reactions were identical to those described in the following section. The primers used were either the *qutA* specific primers and universal primer provided with the Pharmacia T7 sequencing kit.

2.9.3 **Direct sequencing of dsDNA amplified by PCR**

Template for direct sequencing of PCR products was prepared as described (Section 2.7). DNA sequencing was carried out by the chain termination method (Sanger *et al.*, 1977) and products labelled with α^{35}S dATP using the Pharmacia T7 sequencing kit. All reagents used in the sequencing reactions were provided in the Pharmacia T7 sequencing kit.

i) Annealing the primer: A volume of 2µl of primer solution was added to 8µl of the template DNA and heated to 100°C in an Eppendorf for 6min, to denature the template and primer. The tube was then centrifuged briefly to collect condensate and 2µl Annealing buffer added. The tube was again centrifuged to combine the contents and incubated at 42°C for 20min.

ii) Sequencing reactions: The tube was centrifuged briefly to collect any condensate and 3µl labelling mixture, 2µl T7 (2:8 dilution) and 1µl α^{35}S dATP added. The tube centrifuged briefly to mix the contents and held at room temperature for 5min.

iii) Addition of chain terminators: Following the 5min incubation, equal aliquots (4.5µl) of the reaction mixture were added to four tubes each containing 2.5µl of the termination mixes that had been pre-warmed to 42°C. The contents of the tube were mixed by centrifugation and then returned to the 42°C incubator for 5min.

iv) Stopping the reactions: The reactions were stopped by the addition of 5µl stop solution. the samples could now be stored on ice for a few hours or at -20°C overnight until ready to load onto the polyacrylamide gel. To denature the DNA the samples were boiled for 5min immediately prior to loading.

2.9.4 **Preparation of the sequencing gel**

Polyacrylamide/urea gels for the resolution of dideoxy-terminated fragments were normally poured between glass plates of 33 x 41 cm and 33 x 43 cm. Before assembling the plates were cleaned with sterile Q-water and ethanol. The smaller plate was siliconised by application of approximately 2ml of Sigmacote (Sigma). Plates were separated by 0.4 mm thick plastic side spacers and the bottom edge sealed with tape.

Sequencing gel recipe: 37g Urea; 11.5ml Acrylamide 40%; 7.5ml TBE 10x; Made up to 75ml with distilled water.

Acrylamide which was obtained as a 40% solution (Acugel 40: sequencing grade, National Diagnostics) which consisted of acrylamide and bisacrylamide in the ratio 19:1. The mixture was then warmed and stirred to dissolve the urea. The mixture was filtered through Whatmann filter paper then degassed under vacuum. The following were added before the gel was poured. 1ml 10% (w/v) ammonium persulphate; 18µl TEMED.

The polymerising acrylamide was poured between the glass sequencing plates and the top edge of the gel formed using two inverted 0.4mm thick shark tooth combs (BRL). The plates were then clamped together with bulldog clips and polymerisation allowed to proceed for at least 2hr.

When the gel was fully polymerised wells were formed at the top edge using the inverted 0.4mm shark tooth combs and the sequencing gel assembled vertically in BRL electrophoresis sequencing tanks.

Top and bottom reservoirs were filled with 1x TBE buffer and the gel pre-run at 60W constant power for 1hr before loading the sequencing reaction. Prior to loading the gel each well was flushed with 1x TBE buffer to remove accumulated urea.

Sequencing reactions were divided into two equal aliquots and the first aliquot loaded onto the gel. Electrophoresis was carried out at 60W constant power. The second aliquot was then loaded into adjacent wells and electrophoresis carried out as above. The gel running times varied depending upon the separation required, which was dependent upon the distance of any mutation from the sequencing primer.

Once the electrophoresis was complete the plates were then separated and the gel was then fixed for 1hr by emersion in 10% methanol, 10% glacial acetic acid. The gel was then transferred to a sheet of Whatmann 3 MM paper, covered with Saran Wrap and dried under vacuum at 80^{o}C for 2hrs. After removal of the Saran Wrap a Geiger counter was placed on the gel to check for incorporation of the label. The gel was placed in a cassette with x-ray film and stored at -80^{o}C for 2-9 days, dependent upon the level of incorporation.

2.10 Source of Materials

All commonly used reagents and chemicals were of analytical grade unless otherwise stated. Most were supplied by Fisons plc. and Sigma Chemical Co. Ltd., UK Chemicals obtained from other sources are given below
Radionucleotides (Amersham, UK); Acrylamide (National Diagnostics, UK); TEMED (Bio-Rad, USA); HGT agarose (FMC BioProducts, USA); IPTG & X-Gal (Novo Biochem, Denmark); T7 Sequencing kit, T4 DNA ligase & kinase, dNTPs & ddNTPs (Pharmacia Ltd, UK);X-ray film (Amersham, UK/ Fugi, Japan); Taq polymerase (Advanced Biotechnologies, UK); Restriction endonuclease (BRL, UK); *E.coli* strain DH5α (Gibco BRL Life Technologies).

2.11 Containment and Safety

All experimental procedures in this work were done with reference to the Genetic Manipulation Advisory Group's guidelines on safety and containment conditions for such work. All work done fell within the category of Good Microbial Practice. All work carried out was done within the guidelines of the Genetics Department safety book and COSHH recommendations.

Chapter Three

Preliminary mapping of *qutA* mutations in *A.nidulans*

Introduction

The ultimate aim of the work described in this thesis was to identify functional domains of the *qutA* activator protein. This was approached by DNA sequence analysis of twenty seven non-inducible *qutA* mutants and deduction of the resultant changes that render the mutant protein ineffective. Since cloning and total sequence analysis of each *qutA* mutant gene was impractical, a method was required to determine the region of the gene within which each mutation is located.

The work described in this chapter sought to identify regions of the mutant *qutA* genes for DNA sequence analysis. Since the mutations under analysis had been genetically mapped, a successful and reliable alignment of the genetic and physical maps would enable an informed selection of strains for DNA sequence analysis based upon the existing genetic data alone. This may be achieved by mapping the physical locations of a few key mutations.

A method for the identification of single base differences in human DNA sequences, either new mutations or existing polymorphism, has been recently described by Orita *et al.*, (1989). The technique of analysis by single stranded conformational polymorphism (SSCP) is based on two principles. First, the electrophoretic mobility of a ssDNA molecule is dependent upon the shape or conformation of that molecule as well as its length. Second, a single base difference between two otherwise identical ssDNA molecules can yield two types of molecule with altered conformation and this difference may be detected by comparison of their electrophoretic mobility.

The technique is most suitable for the detection of mutations in DNA fragments varying from 50 to 350bp in length. The detection of mutations is also related to the sequence context in which the mutation occurs. The use of overlapping DNA fragments increases the efficiency of detection. To produce such overlapping sequences by PCR would require a large number of primers. More recent reports (Berta *et al.*, 1990; Iwahana *et al.*, 1992) described the use of restriction endonuclease digestion of amplified DNA fragments in SSCP. A large region of DNA was amplified by PCR and cleaved by restriction endonuclease digestion to form optimal size fragments for SSCP analysis. The use of multiple restriction endonuclease digestions was described by Mackay *et al.*, (1993). Different restriction fragments vary the sequence context of any mutation increasing detection efficiency. Additionally this allows a mutation to be located by restriction mapping to a smaller region within the DNA fragment under investigation. Dr R. Dalgleish and Dr

K. Mackay, (University of Leicester, Department of Genetics) provided guidance and assistance in establishing this technique within the laboratory. The essential features of the technique are outlined below.

A specific region of genomic DNA is amplified from control and test sources using synthetic oligonucleotide primers in the polymerase chain reaction (PCR) (Saiki *et al.*, 1988). The primers are designed using the DNA sequence of the gene under study. Restriction endonucleases, selected using the known sequence data, are used to digest the amplified DNA into restriction fragments of a in which alterations in mobility are best observed. The restriction fragments are then denatured before being subjected to gel electrophoresis. Comparison of the mobility of ssDNA fragments amplified from control and test sources can thus physically locate sequence alterations to restriction fragments. To apply this technique it is essential that the DNA sequence of the gene under study is known.

The sequence of the wild-type *qutA* gene has been determined (Beri *et al.*, 1987). Genomic DNA from the wild-type strain (*R153*) provides a control for comparison with amplified DNA from the *qutA* mutant strains. The recipes and protocol for the technique is given in Section 2.8.

3.1 Design of PCR primers for amplification of regions of the *qutA* gene

The analysis of *qutA* mutants by SSCP was begun by Evert von Schothorst. He designed two sets of synthetic oligonucleotide primers (sets 1 and 3) using the 'Primers' program (Lowe *et al.*, 1990) for amplification of *qutA* DNA fragments by PCR. The *qutA* sequence is shown in Appendix 1 and all reference to base position in the sequence are numbered with respect to the translation start codon AUG (bp +1) of the uninterrupted *qutA* open reading frame (ORF).

Selection of regions for SSCP analysis: Primer set 1 was designed using a 1,000bp section of *qutA* (-297 to +703) as reference for the 'Primers' program. The DNA fragment produced by these primers was designated Region 1. This region was selected since it incorporates the coding sequence for the Zinc bi-nuclear cluster putative DNA binding motif (+144 to +228). Mutations in this region may be expected to produce a non-inducible phenotype and the region was therefore likely to encompass at least some of the *qutA* mutations, presenting a potentially productive starting point for the analysis. Additionally, Region 1 includes the translation start codon (+1 to +3) and the pathway specific domain (+228 to +270). Techniques for DNA sequence analysis most readily yield sequence data from 20 to 350 nucleotides 3' of the sequencing primer. Therefore from the possible primers provided by the program, the Region 1 primers were selected to produce an amplified DNA

fragment whose length was such that sequence analysis of the entire fragment was possible using primers at each end of the fragment.

Primer set 2 was constructed by myself at later date, and was designed for analysis of Region 2 between Regions 1 and 3.

Primer set 3, amplifying Region 3, was also designed using a 1,000bp reference section (+700 to +1700). These primers were designed to incorporate the central portion of the gene, which encodes a part of the *QUTA* protein containing no recognised sequence motifs (Figure 1.5), and any mutations mapping to it may identify a novel functional motif. The primers would amplify a large region of the *qutA* gene to increase the likelihood of detecting a mutation. However the size of the fragment means that it would be difficult to sequence the whole region using external primers alone. Therefore the amplified DNA fragment would require either subcloning, or the design of further primers within the amplified Region 3 for sequence analysis of any mutation found to map to the middle of the region.

The 'Primers' program: The 'Primers' program contains a number of parameters which may be varied according to the nature of the target DNA sequence. The G/C content of *qutA* is 51% and was within the standard parameters used by the program 47-53% . The standard parameters for the G/C content selected therefore would not restrict the choice of primers.

The length and sequence of the primer determines its annealing temperature (Ta) and melting temperatures (Tm), where the Ta is approximately 10°C lower than the Tm. A longer primer or higher G/C content raises the Ta to reduce the chance of non-specific primer binding. The Tm recommended by the program for the primers used was 76-83°C and is designed to ensure that the primer remains bound to the template during DNA synthesis at 72°C, but is readily separated in the denaturation step at 94°C. Pairs of primers were selected with a similar Tm to ensure that the annealing temperature of the two primers were as close as possible. Again this would reduce the chance of non-specific amplification.

The program provided a list of possible primer combinations that would amplify fragments of the reference sequence. Some primer sets could be dismissed from consideration as they had the potential to form heteroduplex DNA. Such primers would anneal to each other or themselves and produce a variety of short fragments by amplification of the primer sequence itself.

DNA homology search for potential primer sequences: Primer sequences short listed for use after satisfying the criteria discussed above were then examined in a computer aided search of the fungal DNA sequences in the EMBL database. Primers with a high degree of homology to sequences other

than *qutA* would be capable of binding to such sequences if a low annealing temperature was used in the PCR. Primers showing a degree of homology with any fungal sequence that would provide an annealing temperature close to that calculated for binding to *qutA* were rejected. This procedure was adopted to reduce further the risk of amplification of DNA fragments not from the *qutA* gene.

The PCR primers selected, their location and the length of the amplified fragments produced are as follows: Primer set 1 spans the region -52 to +505, the sequence of the primers are for the sense-strand primer bp-52 to bp-31 5'GTGTCGAAGATGAGCAATGTGC 3' and for the antisense-strand primer bp505 to bp484 5' TTCTTTCTTGGTCCCCTTCCG 3'. Primer set 3 spans the region 726 to 1577: sense-strand sequence is bp726 to 745 5' ACTAACGGATCATCCCACG 3': the antisense-strand sequence is bp1577 to bp1558 5' AGTTGTCGCTCAAGAGGTCC 3'. The sequences given are the actual sequences of the primers and the co-ordinates refer to the sense strand (Appendix 1). Because the primer sequences are incorporated into the amplified fragment, they do not form part of the test sequence itself which are effectively in Region 1 (-30 to 483) and Region 3 (746 to 1557). However any mutation fortuitously located in a primer binding site may affect annealing such that amplification of a dsDNA fragment from that strain may not be possible. This would result in the linear amplification of a ssDNA product, which may not be of sufficient quantity to be visible by gel analysis. The primers were produced by in the primer synthesis laboratory, Department of Biochemistry at the University of Leicester.

Restriction analysis of Regions 1 and 3: A GCG computer aided search for type II restriction endonuclease recognition and cleavage sites was carried out on the two regions of the *qutA* sequence amplified by selected primers (Devereux *et al.*, 1984). The first restriction maps included all known restriction enzyme sites (data not shown). From these maps, enzymes were selected which would digest the amplified DNA into separable fragments in the range 30 to 350bp and thus suitable for SSCP analysis. This list was then compared with the list of enzymes available in the Genetics Department and a selection of those enzymes found in both lists were used in SSCP analysis.

3.2 SSCP analysis of the *qutA* gene Region 1

Conditions of the PCR: It is necessary to optimise the reaction conditions for PCR when using a new set of primers. Conditions are sought that yield the maximum amount amplified DNA as judged by comparison of the intensity of the single band observed in gel electrophoresis with a known amount of reference marker DNA, in this case *Hin*dIII digested λ phage DNA (*Hin*dIII). E. von Schothorst used Primer Set 1 in the first amplification of *qutA* DNA by PCR.

Wild-type *qutA*$^+$ genomic DNA was prepared from conidia yielding DNA of a sufficient quality for use as template in PCR (Section 2.7). The sense and antisense primers of set 1 were prepared. A denaturation temperature of 94°C for 5min was used as this was expected to completely denature *A.nidulans* genomic DNA. DNA synthesis by Taq polymerase has an optimum temperature of 72°C and extension from the 3' hydroxyl (OH) residue of the primers can incorporate over 1,000 nucleotides per min into the new strand. To provide ample scope for the complete synthesis of a 557bp fragment an extension time of 1min was used.

The annealing temperature is dependent upon the sequence and length of the primers being used and is approximately 10°C lower than the melting temperature that is given as part of the data by the Primers program. The lower the annealing temperature the greater the yield of product as primer binding is enhanced. However a low annealing temperature also can allow the primers to anneal to DNA fragments that are related, though not exact sequence matches leading to the amplification of non-specific products. A range of annealing temperatures between 50 and 60°C was examined. The highest yield of *qutA* specific DNA product, a single amplified product of the expected size (560 bp) was obtained using a Ta of 55°C.

The number of cycles of PCR were varied between 30 and 35 with no detectable improvement in the yield of amplified DNA.

The quantity of genomic DNA used as template in the PCR was varied between 1pg and 1µg with maximum yield of amplified DNA being achieved using quantities in excess of 1ng. The quantity of primer used in the reaction was varied between 2 and 30µM. Quantities in excess of 10µM of each primer gave a comparable yield of amplified product.

In summary, the PCR conditions for a 50µl reaction were 10ng genomic DNA template, 10µM of each primer and 30 cycles of PCR. The three steps of the PCR were 5min at 94°C (denaturation), 1.5min at 55°C (primer annealing) and 1min at 72°C (DNA synthesis). The PCR conditions were found to yield approximately 200ng amplified DNA from a single 50µl reaction mixture.

3.2.1 **SSCP analysis of Region 1 by E. von Schothorst**

Genomic DNA was prepared from four mutant strains listed in their order in the genetic map (*qutA423, 552, 303* and *214*). PCR reactions were set up under the conditions described with the addition of 0.5µl ^{32}P dCTP.

Restriction enzyme digestion each of the isotopically labelled amplification product was carried out using the restriction endonucleases *Alu*I, *Hha*I and *Hpa*II (Figure 3.1). Half of the

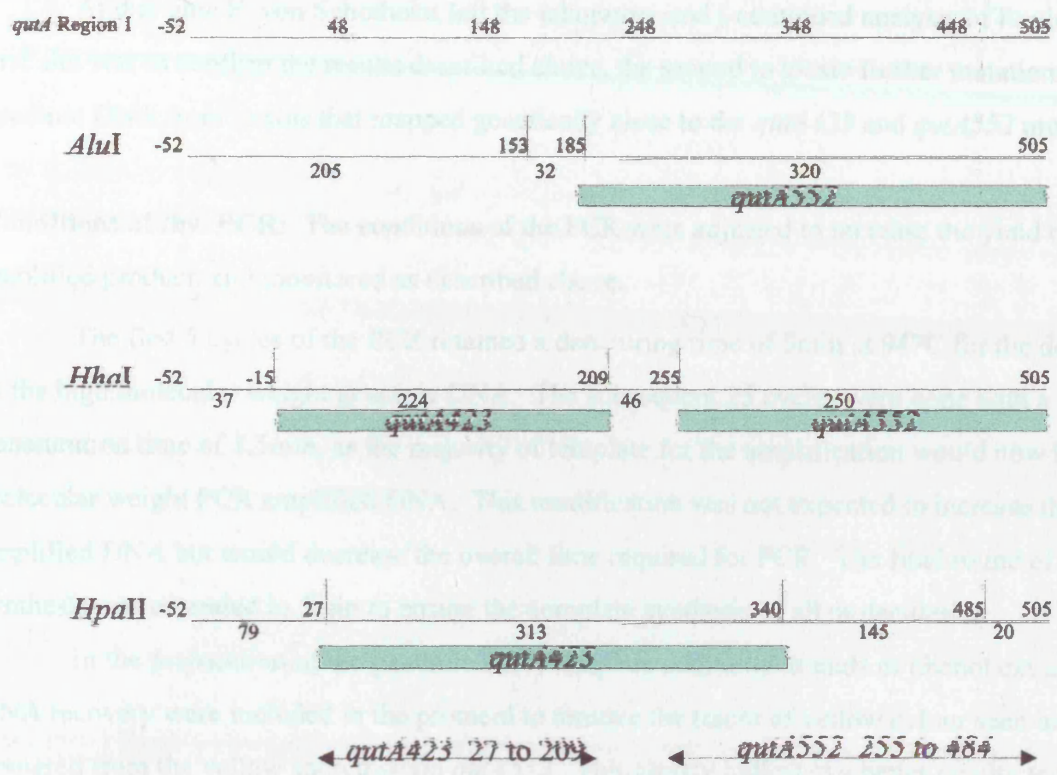
dsDNA from each sample was loaded onto a non-denaturing polyacrylamide gel. This acts as a control showing whether the digestion has produced fragments of the expected number and approximate size. The remaining half of each sample was heat denatured before loading on the gel. Electrophoresis was carried out at 30Watts and 4^{o}C, conditions which allow the formation of the secondary structure within the ssDNA molecules. The gel was dried and exposed to X-ray film for 2-3 days. The resultant autoradiograph is not available for display but provided good quality data for analysis (Similar data obtained by myself are described below in Figures 3.3 and 3.5).

Comparison of the double stranded products from the control and test strains showed the expected restriction fragments. The ssDNA samples identified mutations in the strains *qutA423* and *qutA552* producing ssDNA fragments with altered conformation when compared to the wild-type control. The restriction fragments to which the mutations mapped are shown in Figure 3.1.

The *qutA423* mutation mapped to a 224bp *Hha*I restriction fragment (-15 to +209) and a 313bp *Hpa*II fragment (+27 to +340). The mutation must therefore lie within the region common to both these fragments that is between +27 and +209. The *qutA552* mutation mapped to a 320bp *Alu*I restriction fragment (+185 to +505) and to a 250bp *Hha*I fragment (+255 to +505). Again the mutation must lie in the region common to both fragments that is between +255 and +505. However part of this sequence is the antisense primer (+484 to +505) which does not form part of the test sequence therefore the mutation must lie between +255 to +483. The data show that *qutA423* mutation lies 5' to the *qutA552* mutation. The *qutA423* mutation had been genetically mapped to lie between *qutR* and *qutA552*, that is centromere distal to *qutA552* (Figure 3.2). This result therefore allows a preliminary orientation of the genetic and physical maps which is consistent with the known orientation of the *qutA* ORF. Therefore mutants that genetically map to the left (centromere distal) of *qutA552* lie 5' to this mutation and those mutations mapping to the right (centromere proximal) lie 3' of this site. However a single result for two relatively closely linked mutations cannot be taken to definitively orientate the genetic and physical maps.

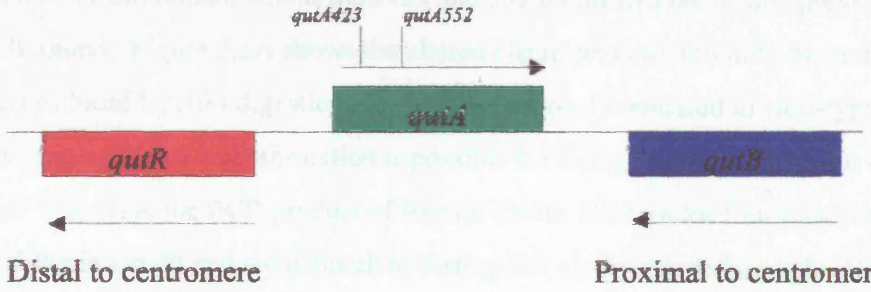
The recombination frequency (RF) observed between *qutA423* and *qutA552* was 70 *qutA*$^{+}$ recombinants per million ascospores tested. The physical distance between the two mutations could be as great as 456bp (+27 to +483) or as little as 46bp (+209 to +255) depending upon the position of the mutant sites within the identified restriction fragment. Such a wide range of possible physical distance between the mutations does not allow the data to provide a reliable guide to the relationship between RF and the actual length of DNA sequence between the mutations.

Figure 3.1 **Restriction map and location of *qutA* mutations**



The figure shows the restriction fragments produced by the digestion of *qutA* Region 1 amplified DNA with the enzymes *Alu*I, *Hha*I and *Hpa*II. Restriction fragments that showed altered electrophoretic mobility are identified by boxes. Minimum sized region containing a mutation is denoted by arrows and bold text. Co-ordinates above each line refer to the *qutA* ORF. The numbers below the line refer to the size of the restriction fragments produced.

Figure 3.2 **Preliminary orientation of the genetic and physical maps**



The figure shows the preliminary orientation of the genetic and physical maps. The orientation shown is consistent with the known orientation of the *qutA* gene Grant *et al.*, (1988). Arrows denote the direction of transcription of the genes.

3.2.2 **Further SSCP analysis of *qutA* Region 1**

At this time E. von Schothorst left the laboratory and I continued analysis of Region 1. The first aim was to confirm the results described above, the second to locate further mutations using genomic DNA from strains that mapped genetically close to the *qutA423* and *qutA552* mutations.

Conditions of the PCR: The conditions of the PCR were adjusted to increase the yield of amplified product, and monitored as described above.

The first 5 cycles of the PCR retained a denaturing time of 5min at 94°C for the denaturation of the high molecular weight genomic DNA. The subsequent 25 cycles were done with a denaturation time of 1.5min, as the majority of template for the amplification would now be low molecular weight PCR amplified DNA. This modification was not expected to increase the yield of amplified DNA but would decrease the overall time required for PCR. The final round of DNA synthesis was extended to 5min to ensure the complete synthesis of all molecules.

In the preparation of the genomic DNA template additional rounds of phenol extraction and DNA recovery were included in the protocol to remove the traces of yellow colour seen in samples prepared from the yellow spored strain *qutA552.* This clearly indicated a better quality template preparation, which was hoped to increase the yield of PCR product. These adjustments to the PCR conditions and template preparation slightly increased the total yield of amplified DNA from approximately 200 to 250ng from a single reaction.

SSCP analysis of Region 1 using strains selected from the genetic map: Further *qutA* mutant strains (*qutA422, 400, 394, 341* and *393*) were chosen for SSCP analysis based upon their close genetic linkage to *qutA423* and *552* (Figure 1.5). Genomic DNA was prepared from each of these *qutA* mutant strains and also the wild-type strain *R153* .

Digestion with the restriction enzymes *Alu*I, *Hpa*II and *Hha*I of the isotopically labelled PCR amplified DNA from mutant strains *qutA423* and *552* confirmed the results previously obtained by E. von Schothorst. Figure 3.3A shows the altered electrophoretic mobility of some 322 base ssDNA fragments produced by *Alu*I digestion of *qutA552* Region 1 compared to wild-type. Figures 3.3A reveals that more than one conformation is possible for a single ssDNA fragment under the same conditions. Digestion the PCR product of Region 1 with *Hha*I yields four fragments two of which are under 50bp in length and are difficult to distinguish on the autoradiograph. Use of *Hae*III to digest the PCR product produced three well separated fragments (247, 160 and 150bp) all of which are readily detected. Since use of *Hae*III improved the resolution of Region 1 it was used instead of *Hha*I in subsequent analysis.

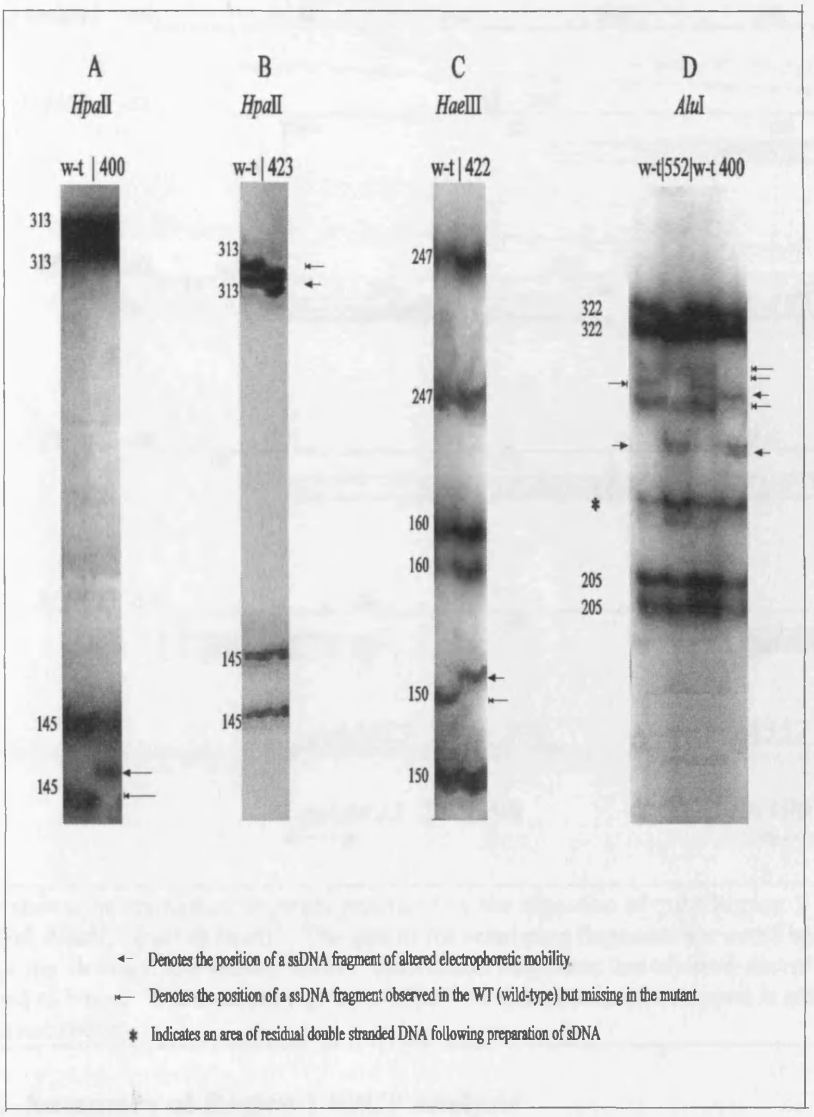
The further analysis of Region 1 located two new mutations. The dsDNA samples in all cases showed the size and number of restriction fragments predicted from the restriction map of the region. Mutant strain *qutA552* which was known to yield fragments of altered mobility when the PCR product is digested with *Hpa*II and *Alu*I was used as a positive control. Digestion with *HpaII* produced three clearly visible bands of 313, 145, 79bp, but the predicted 20bp fragment was not visible. The *Hae*III digest produced the expected three bands of 247, 160 and 150bp.

The ssDNA samples revealed a number of band shifts when compared to the wild-type. The 247bp *Hae*III restriction fragment (+258 to +505) from both strains *qutA552* and *400* showed altered mobility compared to the wild-type. This result marginally improved the location of the *qutA552* mutation from between +255 and +483 to between +258 and +483. The *qutA400* mutation also showed a similar upward shift of the 247 bp *Hae*III fragment locating the *qutA400* mutation to the same fragment. However observation of the *Alu*I digestion of indicates that the *qutA552* and *400* mutations produce different alteration to electrophoretic mobility thus identifying them as different mutations (Figure 3.3A). The *qutA400* mutation was also located to a 145bp (+340 to +483) *Hpa*II fragment (Figure 3.3B), improving the location of the *qutA400* mutation to between bp +340 and +483. The *qutA422* mutation was detected in the same 313bp *Hpa*II fragment as had *qutA423* (Figure 3.3C), and to a 150bp *Hae*III (Figure 3.3D). The *qutA422* mutation was therefore located in the region common to both restriction fragments, namely +27 and +98.

*Alu*I *Hpa*II and *Hae*III digestions of Region 1 PCR products were used in the analysis mutant strains *qutA394, 393* and *341*. However no changes in mobility were identified for any of these strains (data not shown).

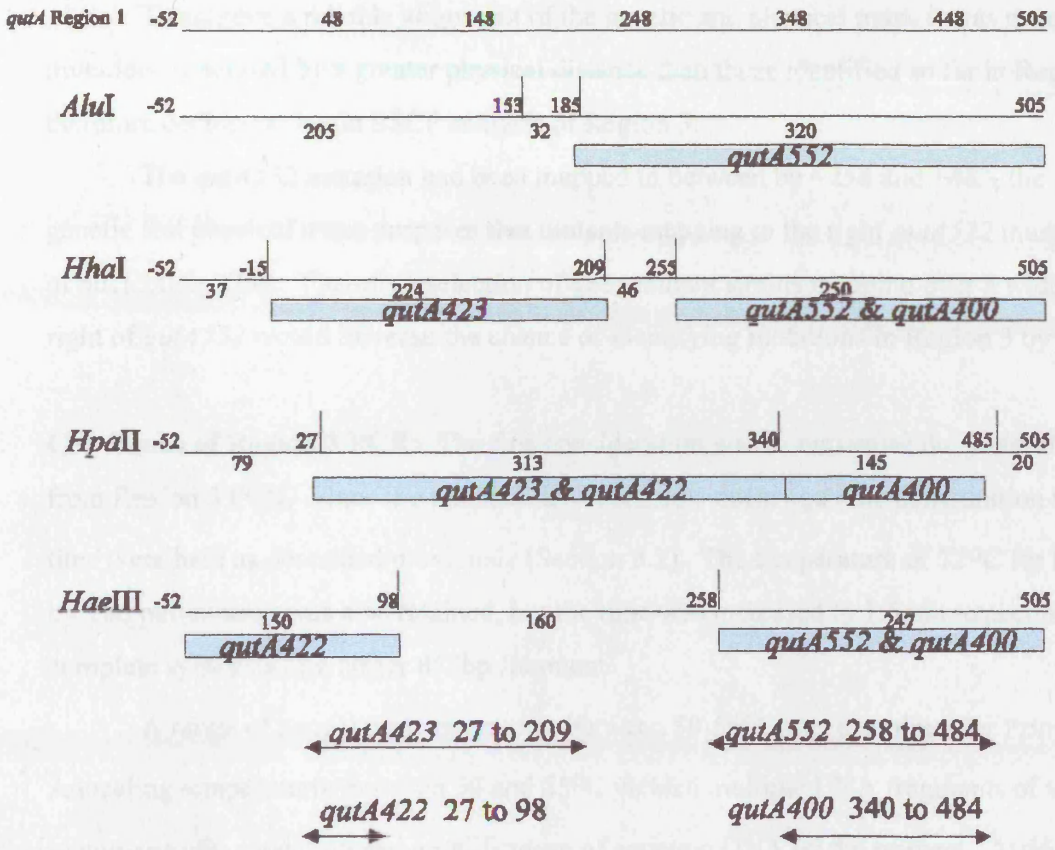
Taken together these additional results brought the total number of mutations mapped by SSCP analysis in Region 1 to four out of nine strains tested (Figure 3.4). The data are summarised in Table 3.1. The location of the two additional mutations physically mapped supports the preliminary orientation of the genetic and physical maps (Figure 3.2). Whilst the mutations *qutA422* and *423* are located in over-lapping fragments as are mutations *qutA400* and *qutA552,* each pair of mutants fall in separate and distinct parts of Region 1. Moreover although *qutA423* and *422* apparently map genetically close together with respect to *qutA552*, the mutation *qutA400* maps much closer to this marker (Appendix 1). Taken together, these data therefore provided strong support to confirm the preliminary orientation of the genetic and physical maps (Figure 3.2). However the physical proximity of the mutations and the overlap of altered restriction fragments again prevents any estimate of the relationship between RF and physical separation.

Figure 3.3 **SSCP analysis of Region 1: Identification of *qutA400*, *423*, *422* and *552* mutations**



The figure shows the moboility shifts observed in ssDNA fragments during SSCP analysis compared to adjacent wild-type (w-t) samples. A) shows a downward shift of 1 of 5 alternate conformations seen with the 313 base ssDNA fragments produced by *Alu*I digestion of Region 1 DNA amplified from *qutA552*. In addition a downward shift of 2 and disappearence of 1 of 5 alternate conformations is observed seen with the 313 base ssDNA fragments produced by *Alu*I digestion of Region 1 DNA amplified from *qutA400*. Thus showing that whilst *qutA400* and *552* lie in the same restriction fragment they are clearly different mutations. B) shows an upward shift of the lower 145 base ssDNA fragment produced by *Hpa*II digestion of Region 1 DNA amplified from *qutA400*. C) shows a downward shift of the upper 313 base ssDNA fragment produced by *Hpa*II digestion of Region 1 DNA amplified from *qutA423*. D) shows an upward shift of the upper 313 base ssDNA fragment produced by *Hae*III digestion of Region 1 DNA amplified from *qutA422*.

Figure 3.4 **Restriction map of Region 1 and location of *qutA* mutations identified by SSCP**



The figure shows the restriction fragment produced by the digestion of *qutA* Region 1 amplified DNA with the enzymes *Alu*I, *Hha*II, *Hpa*II & *Hae*III. The size of the restriction fragments are noted below each line and the co-ordinates of the cleavage site shown above. Restriction fragments that showed altered electrophoretic mobility are displayed as boxes. The smallest region to which the mutation can be mapped is noted by arrows and the co-ordinates given above.

Table 3.1 **Summary of Region 1 SSCP analysis**

Mutant strain tested	Location of mutation	Genetic map distance from marker strain *qutA552*
qutA423	27 to 209	left 70
qutA422	27 to 98	left 63
qutA400	340 to 483	left 19
qutA552	258 to 483	Reference strain
qutA394	3' to 258	left or right 3
qutA341	3' to 258	left or right 6
qutA393	3' to 258	right 10
qutA303	3' to 258	right 168
qutA214	3' to 258	right 286

The table summarises the physical and genetic data for those strains used in SSCP analysis of Region 1.
The physically mapped location of the first four strains is given. In addition the predicted location of the remaining five strains tested is provided. The genetic distance is given in *qutA*$^+$ recombinants per million ascospores plated (see also Figure 1.6 and Appendix 2). Left refers to mutations centromere distal of *qutA552*, that is between *qutA552* and the *qutR* marker. Right refers to mutations centromere proximal to *qutA552*.

3.3 SSCP analysis of the *qutA* gene Region 3

To achieve a reliable alignment of the genetic and physical maps it was necessary to identify mutations separated by a greater physical distance than those identified so far in Region 1. It was therefore decided to begin SSCP analysis of Region 3.

The *qutA552* mutation had been mapped to between bp +258 and +483, the alignment of the genetic and physical maps proposes that mutants mapping to the right *qutA522* must therefore lie 3' of nucleotide +258. Therefore selection of *qutA* mutant strains mapping over a wide range to the right of *qutA552* would increase the chance of identifying mutations in Region 3 by SSCP analysis.

Conditions of Region 3 PCR: The first consideration was to maximise the yield of amplified DNA from Region 3 PCR. Since the template DNA remains unchanged the denaturation temperature and time were held as described previously (Section 3.2). The temperature of 72°C for DNA synthesis by Taq polymerase was also retained, but the time was increased to 1.5min to accommodate the complete synthesis of a larger 852bp fragment.

A range of annealing temperatures between 50-56°C was examined for Primer set 3. Annealing temperatures between 50 and 55°C yielded multiple DNA fragments of various sizes due to non-specific annealing and amplification of genomic DNA by the primers. At 56°C the PCR yielded a single DNA fragment of the expected size.

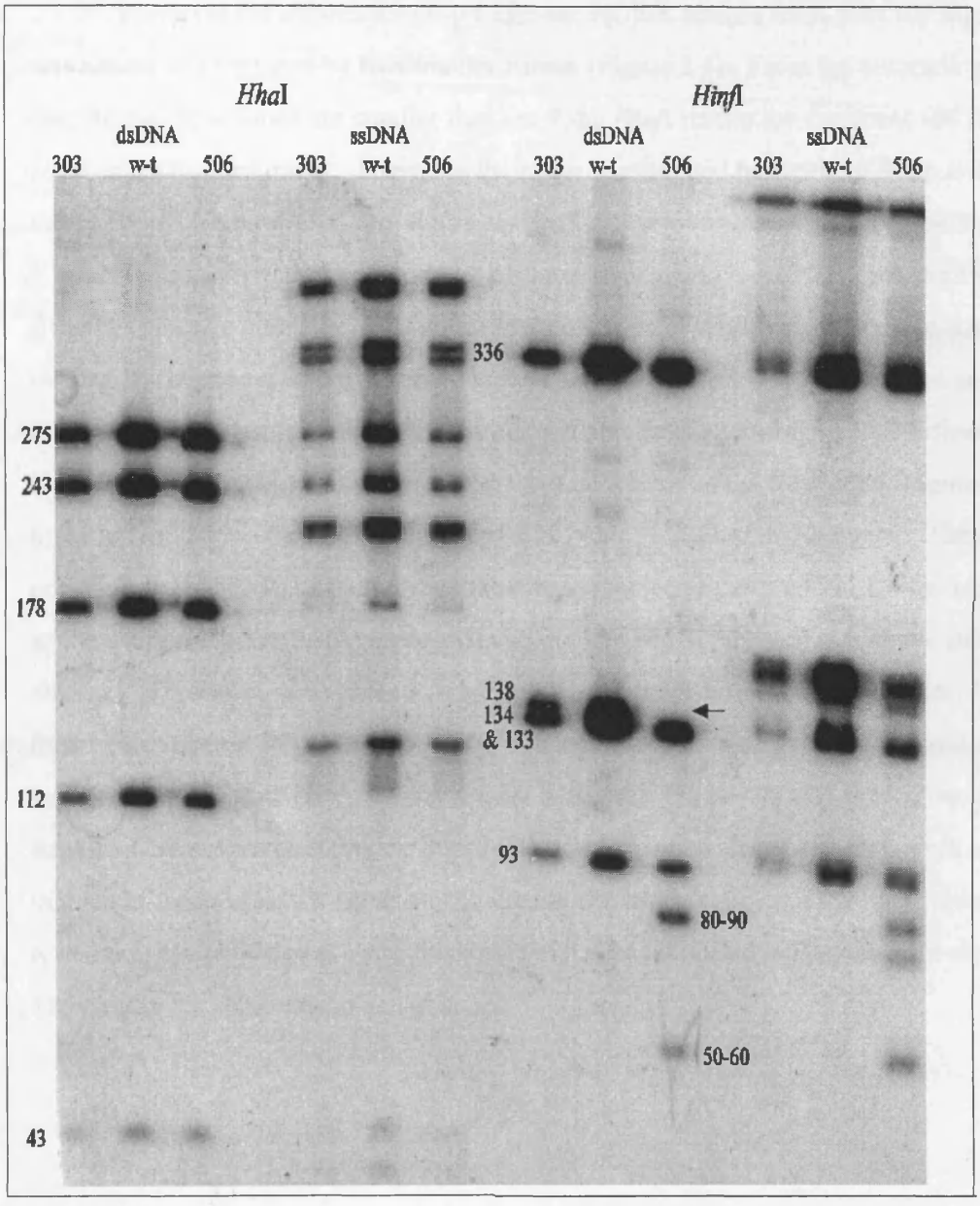
The quantity of template (10ng) and primer (10μM) used in the amplification of Region 3 were varied and found to be optimal at the levels previously described for Region 1. PCR with Region 3 primers and wild-type genomic DNA as template also yielded approximately 250ng amplified product from a 50μl reaction.

SSCP analysis of *qutA* Region 3: A preliminary set of four mutant strains (*qutA4, 506, 303* and *214*) were chosen and genomic DNA was prepared for use as template in the PCR.

The isotopically labelled amplification products were digested using each of the restriction endonucleases *Alu*I, *Ava*II, *Hha*I *Hinf*I or *Sty*I. Samples wild-type and mutant amplified DNA were analysed on a non-denaturing polyacrylamide gel as before.

Figure 3.5 shows an autoradiograph of the PCR amplified DNA from for two mutant strains *qutA303* and *506* and the wild-type control digested with the enzymes *Hha*I and *Hinf*I. The dsDNA samples from the wild-type and mutant strains *qutA303, 506* confirmed that digestion with *Hha*I produced five fragments whose estimated size agree with the predicted restriction fragments of 275, 243, 178, 112 and 43bp. The dsDNA samples from the wild-type and mutant strain *qutA303*

Figure 3.5 **SSCP analysis of Region 3: identification of the *qutA506* mutation**

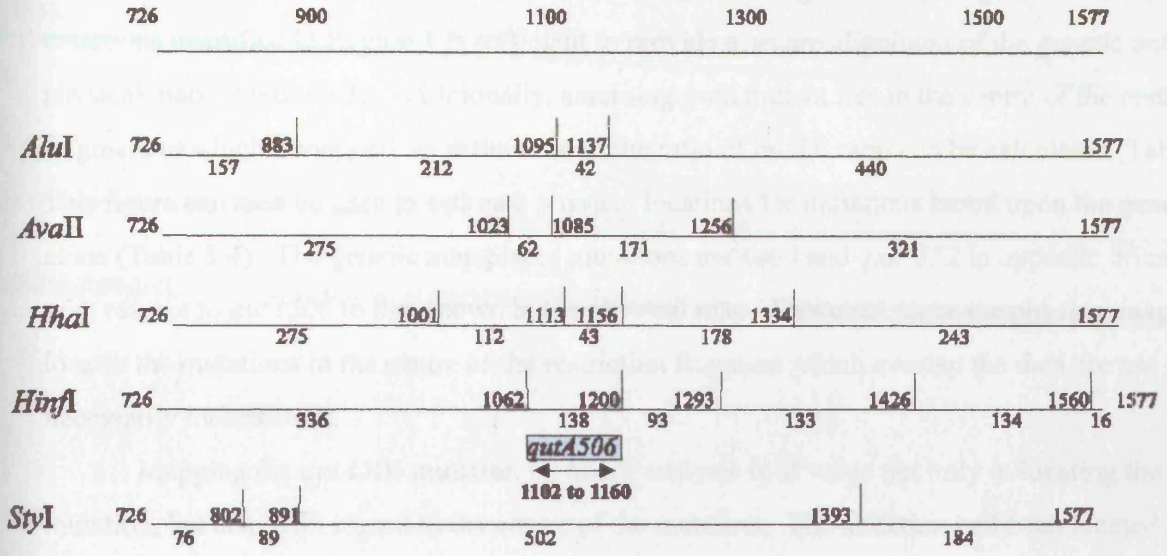


The figure shows the auotradiograph produced by SSCP analysis of Region 3 PCR from the mutant strains *qutA303*, *506* and the wild-type digested with *Hha*I and *Hinf*I. The arrow indicates the missing 138bp in the *qutA506 Hinf*I. The two additional bands appearing in the gel are labelled 80-90 and 50-60bp. This reflects the creation of a *qutA506* mutation that creates a new *Hinf*I restriction site.

confirmed that digestion with *Hinf*I also produced the expected five visible fragments of 336, 138, 134, 133, and 93bp; the predicted 16bp band is not visible.

However the expected 138bp fragment dsDNA sample from *qutA506* digested with *Hinf*I was absent and replaced by two smaller bands. (Figure 3.5). From the autoradiograph we can see that the two new bands are smaller than the 93bp *Hinf*I restriction fragment and larger than the 43bp *Hha*I restriction fragment. Therefore their size is estimated between 80-90bp and 50-60bp (mean total 140bp). The *qutA506* mutation has therefore created a new *Hinf*I recognition and cleavage site 5' GANTC 3'. Inspection and analysis of the region +1062 to +1200 in the *qutA* sequence revealed five sites where a single base substitution would form a *Hinf*I recognition site that would yield restriction fragments of the observed size. This result may assist the sequence analysis of the mutation. Consideration of the size of the two new restriction fragments confines the possible location of the mutation to at least 40bp from either end of the restriction fragment. The *qutA506* mutation must therefore lie between bp +1102 and +1160 in the *qutA* gene. The ssDNA samples revealed no mobility shift other than the disappearance of the *qutA506* 138bp band and the addition of the smaller bands. SSCP analysis using the five restriction enzymes above failed to identify any changes in mobility of fragments amplified from mutant strains *qutA4*, *303* and *214*. A further seven mutant strains (*qutA633, 324, 361, 389, 390* and *636*) that map genetically close to *qutA506* were selected and genomic DNA prepared from each. SSCP analysis was carried out on these seven strains. Five separate restriction digests *Alu*I, *Ava*II, *Hha*I, *Hinf*I and *Sty*I were used in the analysis of each of these strains without any change in electrophoretic mobility being detected. The restriction map of Region 3 and the single mutation identified in Region 3 are shown in Figure 3.6. The results are summarised in Table 3.2.

Figure 3.6 **Restriction map of Region 3 and location of *qutA506* identified by SSCP**



The figure shows the restriction fragment produced by the digestion of *qutA* Region 3 amplified DNA with the enzymes *Alu*I, *Ava*I, *Hha*II, *Hinf*I & *Sty*I. The size of the restriction fragments are noted below each line and the co-ordinates of the cleavage site shown above. Restriction fragments that showed altered electrophoretic mobility are displayed as boxes. The smallest region to which the mutation can be mapped is noted by arrows and the co-ordinates given above.

Table 3.2 **Summary of Region 3 SSCP analysis**

Mutant strains	Location of mutation	Genetic distance from marker strain qutA506
qutA4	5' to 1160	left 99
qutA633	5' to 1160	left 61
qutA324	5' to 1160	left or right 3
qutA361	5' to 1160	left or right 2
qutA506	1102 to 1160	Reference strain
qutA390	3' to 1102	left or right 1
qutA389	3' to 1102	left or right 2
qutA636	3' to 1102	right 16
qutA303	3' to 1102	right 17
qutA214	3' to 1102	right 54

The table summarises the physical and genetic data for those strains used in SSCP analysis of Region 3.
The physically mapped location of *qutA506* is given. In addition the predicted location of the remaining ten strains tested is provided. The genetic distance is given in *qutA*$^+$ recombinants per million ascospores plated (see also Figure 1.6 and Appendix 2). Left refers to mutation centromere distal of *qutA506*, that is between *qutA506* and the *qutR* marker. Right refers to mutations centromere proximal to *qutA506*.

Discussion

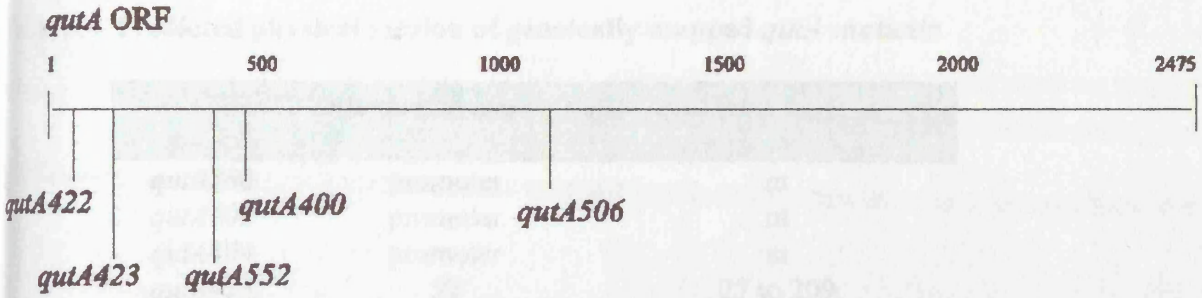
The single mutation identified by SSCP analysis in Region 3 taken together with those mutations identified in Region 1 is sufficient to provide a secure alignment of the genetic and physical maps (Figure 3.7). Additionally, assuming each mutant lies in the centre of the restriction fragment to which it mapped, an estimation of the ratio of bp:RF ratio can be calculated (Table 3.3). This figure can then be used to estimate physical locations for mutations based upon the genetic data alone (Table 3.4). The genetic map places mutations *qutA400* and *qutA552* in opposite orientations with respect to *qutA506* to that shown in the physical map. However, since the physical map simply locates the mutations in the centre of the restriction fragment which overlap the data are not necessarily inconsistent.

Mapping the *qutA506* mutation by SSCP analysis is of value not only in locating the mutation, but also with regard to the nature of the mutation. The mutation had been located to the region 1102 to 1160 and found to create a new *Hinf*I restriction site (5' GANTC 3'). Assuming that the mutation affected only a single base, there are five substitution mutations and eight addition or deletion mutations which have the potential to form *Hinf*I sites in this 58 bp sequence. The high density of potential sites of mutation will require careful DNA sequence analysis to identify the actual mutation.

The predicted physical locations of *qutA* mutations, (Table 3.4) places ten (*qut423, 422, 400, 394, 552, 341, 393, 386, 4* and *384*) in Region 1 and fourteen (*qutA385, 399, 633, 379, 324, 361, 506, 389, 390, 636, 303, 444, 382* and *214*) in Region 3. While four of seven mutations tested in Region 1 had been located by SSCP analysis, only one of eight tested in Region 3 was detected. A negative result in SSCP analysis implies that, under the conditions employed, restriction fragments have the same electrophoretic mobility though not necessarily the same sequence. Therefore the conclusion cannot be drawn that a mutation is not present, rather that it cannot be detected by the technique.

Adjustments to parameters of SSCP analysis could have been undertaken to identify other mutations and thus further secure the genetic and physical alignment. Reports have been made of improved mutation identification using different running temperatures, decreasing the glycerol percentages from 10% to 0% (Spinardi *et al.*, 1991) and high percentage acrylamide:bis-acrylamide gels (Savov *et al.*, 1992). A more radical alteration to the protocol could have been undertaken. Incorporating phage promoter sequences within the PCR primers allows an additional translation stage to produce RNA copies of amplified DNA fragments (rSSCP). RNA copies yield more readily detected conformational changes from single base differences (Sarkar *et al.*, 1992). However it was felt that the alignment achieved with the data thus far was sufficient to proceed with DNA sequence

Figure 3.7 **Alignment of the genetic and physical maps of the *qutA* activator in *A.nidulans***



The figure shows the five mutations identified by SSCP analysis and their relationship to the physical map of the *qutA* gene. The mutations are shown in the centre of the restriction fragment to which they map. The co-ordinates above the line refer to the nucleotide sequence of the *qutA* ORF.

Table 3.3 **Genetic and Physical distances between mutations identified by SSCP analysis**

Mutant strains	RF	Estimated physical distance between mutations (bp)	Ratio bp/RF
qutA423 to *qutA552*	70	264	3.6
qutA422 to *qutA552*	63	319	4.9
qutA400 to *qutA552*	19	Overlap	Not applicable
qutA423 to *qutA506*	Not done	964	Not applicable
qutA422 to *qutA506*	232	1068	4.6
qutA400 to *qutA506*	174	718	4.1
qutA552 to *qutA506*	181	760	4.2
			Average = 4.3

The table shows the RF distance between mutations (see Appendix 2). The physical distance between mutations is estimated assuming that mutations lie in the centre of the restriction fragment to which they map. The average ratio between bp and RF can be used to predict the location of unidentified mutations using the physical data from mapped mutations and the genetic data alone.

Table 3.4 **Predicted physical locaion of genetcally mapped *qutA* mutants**

Mutant strain	Predicted location in ORF	SSCP data
qutA380	promoter	nt
qutA402	promoter	nt
qutA404	promoter	nt
qutA423	52	27 to 209
qutA422	82	27 to 98
qutA400	271	340 to 484
qutA394	340	x
qutA552	353	255 to 484
qutA341	379	x
qutA393	396	x
qutA386	426	nt
qutA4	452	nt
qutA384	474	nt
qutA399	843	nt
qutA385	869	nt
qutA633	869	x
qutA379	878	nt
qutA324	1119	x
qutA361	1123	x
qutA506	1131	1102 to 1160
qutA390	1136	x
qutA389	1140	x
qutA303	1196	x
qutA636	1200	x
qutA444	1213	nt
qutA382	1273	nt
qutA214	1364	x

The table shows the predicted physical location of each *qutA* mutation estimated using the centre of the *qutA506* restriction fragment as a reference. The genetic data used was taken from Appendix 2 and the bp:RF ration was taken from Table 3.3. nt represent strains not tested by SSCP analysis. x represents no alteration in electrophoretic mobility detected.

analysis of the mutations. DNA sequence analysis would allow further improvement of the alignment and provide the primary data required for the characterisation of the *qutA* mutations.

DNA sequence analysis of the region of the *qutA* gene predicted by the genetic data, to contain a particular mutation will establish if SSCP analysis had simply not detected that mutation. If this is the case the alignment can then be used to identify further regions of the *qutA* gene for sequence analysis. A progressive improvement in the alignment of the genetic and physical maps may be expected as mutations are mapped by DNA sequence analysis. Mutations identified in Regions 1 or 3, by sequence analysis but not identified by SSCP analysis can in turn yield further information about the limitations of SSCP.

Taken overall two major predictions emerge from the provisional alignment and physical maps (Figure 3.8). First, and perhaps not unexpectedly, certain mutations fall in the promoter region. One is recessive (*qutA380*) and two (*qutA404* and *402*) are semi-dominant. Presumably these mutations affect levels of *qutA* protein produced. This may require re-interpretation of the semi-dominant phenotype as described in Section 1.3.3.

Second, and **most unexpectedly** no non-inducible mutations fall in the 3' carboxyl half of *qutA* gene. This at once raises questions about the function of this part of the protein. Is this region of the protein superfluous or does it have a function different from activation of transcription identified for the carboxyl terminus? It is clear that sequence analysis of the *qutA* mutations will yield interesting new data about the activator protein.

Chapter Four

Part I

Approaches to sequencing regions of the *qutA* gene

Introduction

The aim of the work described in this chapter was to develop a rapid and reliable method by which selected regions of the *qutA* gene can be sequenced from mutant strains.

Analysis by SSCP has provided a method of locating mutations to regions of the gene. Four mutations (*qutA422*, *423*, *400* and *552*) have been located to defined restriction fragments within Region 1 and one (*qutA506*) in Region 3. Data from the analysis of these mutations enables the provisional location of further mutations based upon their position in the genetic map. The process of amplifying regions of the *qutA* gene by PCR for SSCP analysis also yields amplified DNA that could be used as template in sequencing reactions.

A number of procedures were followed in attempting to sequence Regions 1 and 3. The first two proved unproductive, but the third gave sequence data that was instrumental in determining the successful strategy described in Chapter 5.

Sequence data from clones amplified by PCR as the result of a Region 1 primer binding to sequences other than the *qutA* gene proved of interest, and is described in Part II of this chapter.

4.1 DNA sequence analysis using ssDNA template produced by PCR

This method proposes the production of ssDNA for use as sequencing template by a two step PCR technique described by Gyllensten *et al.*, (1988) and Bachmann *et al.*,(1990). The first step is the amplification of target DNA sequence from genomic DNA using two primers in a standard PCR. In the second step the amplified dsDNA is used as template in a PCR reaction that contains only one primer, thus producing defined ssDNA for use as template in the sequencing reactions. A more detailed account of the following procedure is given in Section 2.9.1.

Step 1. The production of dsDNA amplified by PCR: Genomic DNA from the wild-type strain *R153* was used as template for amplification of Region 1 of the $qutA^+$ gene. Primer set 1 was used in the reaction under the conditions described in Section 3.1. A single 50ml reaction yielded 200ng amplified DNA. To provide sufficient template for the second step, five parallel PCR reactions were

done, the products combined and the amplified fragment separated from excess primer and un-incorporated nucleotides by gel electrophoresis. The amplified DNA was recovered from the gel yielding approximately 900ng amplified Region 1 dsDNA as judged by the band intensity in gel electrophoresis.

Step 2. The production of ssDNA amplified by PCR: PCR reactions using two primers amplify the target sequence exponentially with the number of cycles employed, since the amount of template is doubled each cycle. PCR using a single primer amplifies the target sequence linearly with the number of cycles, because the amount of template remains constant throughout the reaction. Step 2 PCR using 200ng dsDNA from step 1 contains 100ng ssDNA template. This template amplified in 16 cycles is expected to produce 1.6μg of ssDNA. In order to conserve material no analytical gel was done and I sought to recover this ssDNA produced by a selective precipitation.

Step 3 DNA sequencing reactions: The product from step 2 was resuspended in 8μl of distilled water and used as template in sequencing reactions with the alternative primer to that which produced the template. Solutions for the sequencing reactions were prepared following recipes based upon the Pharmacia T7 sequencing kit. Sequencing reactions were carried out as described in Section 2.9.1, with M13 template and the associated universal primer provided in the sequencing kit used as a positive control in the experiment. The resultant autoradiograph showed a sequence ladder for the M13 template but no incorporation of radiolabelled nucleotides in the experimental samples (data not shown).

The dsDNA product from step 1 had been observed in gel electrophoresis and the control sequencing reaction had worked. Therefore the failure of the experimental sample to yield any signal in the sequencing reactions lay in either the amount or quality of the ssDNA template obtained in the step 2 PCR, or the use of the PCR primers as sequencing primers.

Adjustments to the procedure: As a general rule primers do not bind efficiently at the ends of fragments. Both the step 2 PCR and the sequencing reactions rely on the primers binding to the end of the amplified fragment. If primer binding was inefficient the result would be low yields of step 2 PCR and thus insufficient template for the sequencing reactions.

The first stage in assessing the problem was to attempt to assay the yield from step 2 PCR. Spectrophotometry could not be used to determine the amount of ssDNA due to interference from the excess primer and unincorporated nucleotides. In an attempt to visualise the ssDNA, the yield from ten PCR reactions were combined and used in gel electrophoresis. However no band was observed. Since ssDNA does not intercalate ethidium bromide as efficiently as dsDNA it was decided to determine the quantity of ssDNA necessary for visualisation with ethidium bromide in gel electrophoresis. Various quantities of λ *Hin*dIII digested DNA were denatured by boiling for six minutes and subjected to electrophoresis. At least 100ng of the 560 base ssDNA fragment were

required for a discernible signal. It was therefore not possible to monitor the production of template in step 2 unless the yield was increased to 100ng. Similarly adjustments to the reaction conditions could not be monitored effectively.

Summary: It seems likely that the underlying difficulty experienced may have been inefficient annealing of primers at the ends of template fragments. This would apply to both the sequencing reactions and step 2 PCR, where the conditions employed may have exacerbated the difficulty. It might have been an advantage to increase the period for DNA synthesis to 5min for the last few cycles but this point was not recognised at the time. Due to the difficulties encountered with this relatively new sequencing technique it was decided to adopt the more traditional approach of cloning the PCR products prior to sequencing. This would at least answer the question, whether the problem lay in this particular technique or in some fundamental error that I was repeating.

4.2 Cloning PCR amplified DNA into the vector phage M13mp18

Cloning segments of DNA amplified by PCR into the circular molecule of a sequencing vector has three major advantages over the direct sequencing a linear of PCR amplified template. First, the sequencing primer binds to an internal vector sequence close to the target sequence. Second, the cloned fragment can be amplified to provide sufficient quantities of template for sequencing. Finally, segments may be deleted from a large cloned fragment, to bring central regions close to the sequencing primer.

The vector M13mp18 was used since it was the standard sequencing vector used in Dr. A. R. Hawkins laboratory where this work was carried out. *Escherichia coli* infected with phage M13 and plated on agar produces faint 'plaques' which represent zones of inhibited growth. Additionally the vector contains the *E.coli lacZ* gene including a proximal multiple cloning site (MCS). Insertion of a cloned sequence at the MCS disrupts synthesis of the α-peptide of β-galactosidase and the plaque is colourless on X-gal plates rather than blue due to the metabolism of the substrate to a strongly coloured compound. The host strain of *E.coli* is deleted for the native *lacZ* gene, but contains an F' plasmid encoding the β-peptide of the enzyme.

The cloning and subcloning strategy: It was decided to begin the work with the Region 3 fragment to provide experience in both the cloning and subcloning strategies described below. Briefly, Region 3 DNA amplified by PCR will be cloned into the *Sma*I site in the vector by 'blunt ended' ligation. The fragment may be inserted in either orientation. The DNA from mature phage is single stranded, therefore clones in both orientations are required to sequence both the Region 3 sense and antisense strands from the phage primer.

PCR amplification often results in the addition of a non-templated nucleotide at the 3' end of the synthesised strand making the fragments 'ragged' which would prevent ligation into the flush ended *Sma*I site. The amplified DNA was incubated with T7 polymerase to 'fill-in' the single stranded sequence (Section 2.6). Additionally the primers incorporated into the synthesised strands lack the 5' terminal phosphate group required for the ligation reaction. Incubation of the amplified fragment with T7 kinase provides the phosphate group for ligation (Section 2.6).

Further subcloning of the Region 3 DNA makes use of a *Hin*dIII restriction site which lies at one end of the fragment in the vector MCS, and a second *Hin*dIII restriction site at bp1135 approximately in the centre of Region 3. Digestion of the recombinant clone with *Hin*dIII would produce two fragments, a large fragment containing the vector sequence and part of Region 3 and a small fragment consisting of the remainder of Region 3 and the vector MCS sequence between the *Sma*I and *Hin*dIII sites. These two fragments can be separated by gel electrophoresis. The larger fragment can then be recovered and re-ligated to produce a subclone of Region 3. The central portion of Region 3 is brought close to the universal primer site in the resultant subclone. Since the Region 3 fragment is expected to be cloned in both orientations with respect to the vector MCS, the same subcloning strategy applied to a clone in the opposite orientation will bring the other half of the insert close to the universal primer for sequencing.

4.2.1 **Cloning Region 3 DNA**

Preparation of the vector: A sample of the dsDNA replicative form of the phage M13mp18 vector (provided by Dr. A. R. Hawkins) was incubated with *SmaI* endonuclease at 30°C for 1.5hr. The linearised vector was treated with Bovine Alkaline Phosphatase to remove the 5' phosphate group from the vector and thus prevent its re-ligation (Section 2.6). The enzymes were then removed from the sample by phenol extraction and the DNA recovered by ethanol precipitation. The sample was resuspended in 100µl distilled water and 10µl used in gel electrophoresis to monitor the concentration of vector DNA available. The procedure yielded 1µg of linear M13mp18 for use in ligation reactions, and the vector DNA was diluted to 5ng/ml.

Preparation of the Region 3 DNA: Genomic DNA from the wild-type (*R153*) and mutant strain *qutA506* was prepared from mycelium. This method produces genomic DNA of a greater purity than the rapid 'mini-prep' from conidia previously employed. It was hoped that the greater purity of this genomic DNA used as template in the PCR would improve the yield of amplified product. However visual comparison of PCR products amplified using genomic DNA from the two different methods showed no difference in the yield.

Genomic DNA from the mutant strain *qutA506* was used in the preparation of Region 3 PCR amplified DNA using the conditions previously established (Section 3.2). The PCR products were

separated by gel electrophoresis and the amplified fragment recovered yielding 200ng of Region 3 DNA. The PCR amplified DNA was treated with T7 polymerase (Section 2.6). The enzyme was removed by phenol extraction and the DNA recovered by ethanol precipitation. The 'blunt ended' PCR product was treated with T7 kinase to provide 5' phosphate groups (Section 2.6), another phenol extraction done to remove the enzyme and the DNA recovered by ethanol precipitation. In order to conserve material no analytical gel was done. The DNA for cloning was finally resuspended in a final volume of 20µl water, the concentration of the insert DNA was assumed to be between 100-150ng in 20µl.

Ligation and transformation: Each set of ligation reactions contained a constant amount of vector DNA, while the volume of Region 3 DNA was varied in the ratio of vector to insert molecules in the range 1:0.5, 1:1, 1:2 for each ligation reaction (Section 2.6). Three experimental and three controls ligations were set up. Control 1 contained no DNA, Control 2 contained linearised vector DNA not treated with phosphatase and Control 3 circular vector.

The products of the ligation reactions were used to transform competent cells of *E.coli* strain JM103 (Section 2.5). The sample containing no added DNA (Control 1) yielded no phage plaques confirming that the JM103 cells, ligation solutions, transformation solutions and growth media were free from phage contaminants. The sample containing the *Sma*I digested vector DNA (Control 2) was a positive control for the ligation reactions, and the blue plaques represent re-ligation of the vector. The sample with circular vector (Control 3) yielded a large number of blue plaques showing that the transformation procedure was successful. The experimental samples each produced a number of white plaques (recombinant phage) and blue plaques. The number of blue plaques from each test plate and Control 2 were roughly equivalent reflecting the number of vector molecules that had not been linearised by *Sma*I, or were not dephosphorylated and thus able to re-ligate. The number of recombinant phage increased with the quantity of Region 3 DNA used and thus the optimal conditions for the ligation reaction were those that yielded the highest number of recombinant clones with the least amount of Region 3 DNA. The volume of Region 3 DNA was doubled between experimental samples 1 and 2 but the number of plaques formed increased over 5 fold. The volume of Region 3 DNA was again doubled between test samples 2 and 3 but was this time accompanied by only a small increase in the number of plaques formed.

Selection and analysis of recombinant clones: Phage particles were recovered from twelve well isolated white plaques and one blue plaque from Control plate 3 and used to prepare phage ds and ss DNA (Sections 2.5.4). The blue plaque provided a source of vector DNA as a control. The dsDNA samples were stored at -20^{o}C until required.

To determine the orientation of the Region 3 DNA in the recombinant clones the procedure of 'T-tracking' was carried out with the ssDNA samples. This procedure involves sequencing reactions

using only the analogue ddTTP to terminate DNA synthesis and therefore requires only one quarter of the materials used in normal sequencing reactions. Comparison of the 'T-tracks' from each clone allows determination of the orientation of the clones with respect to each other. Comparison of the 'T-tracks' with the known DNA sequence allows orientation of the cloned sequence with respect to the *qutA* ORF. Eleven of the twelve recombinant clones were orientated such that the sense strand is available for sequence analysis. One clone was recovered in the opposite orientation, where the antisense strand is available for sequence analysis.

Subcloning the Region 3 DNA from M13mp18: At this time it became necessary to return to Leicester, when samples of vector DNA and the dsDNA prepared from each recombinant clone were retained as the starting materials for the subcloning exercise.

The first step was to amplify recombinant phage dsDNA. The sample of dsDNA from one clone of each orientation A or B were transformed into *E.coli* strain JM103, yielding only white plaques as expected. However the transformation efficiency was low yielding only eleven plaques from the sense strand clone and seven from the antisense strand clone. Phage were recovered from each of the plaques and used to inoculate cultures of strain JM103 for bulk preparation of ss and dsDNA. The freshly prepared cloned dsDNA was incubated with *Hin*dIII and the products separated by gel electrophoresis. The gel revealed that some recombinant clones had not been digested, others were digested but only produced a single fragment and some produced two fragments neither of which were of the expected size.

It appears most likely that the infected cells of JM103 had been incubated for too long a time, resulting in deletion of sequences from recombinant phage. Unfortunately all of the single original recombinant phage preparation in the antisense orientation had been committed to amplification and thus was lost. Repetition of the whole subcloning procedure was thought to be unproductive.

4.3 Cloning target sequences amplified by PCR into the vector pUC18

The sequencing vectors mainly used in the Department of Genetics are the pUC based plasmid vectors, which have a number of advantages over the phage M13 vectors. The protocols for the preparation of plasmid DNA are simpler and the plasmid is stable and less prone to deletion of inserted sequences than the phage vector. Additionally, since the plasmid sequencing template is double stranded sequence data can be obtained from both strands of a single clone by use of the appropriate primer. Therefore clones in both orientations are unnecessary. Finally, as the pUC vector system is commonly used in the department direct technical support was available should any problem be encountered.

The bacterial strain DH5α, which is resistant to nalidixic acid, was used as host for the propagation of pUC, which in turn confers ampicillin resistance. Thus, inclusion of ampicillin and nalidixic acid in the growth medium allows a double selection for DH5α containing the plasmid. The pUC vector MCS is also located at the start of the *lacZ* gene and selection of recombinant plasmids on X-gal medium is available (Section 4.2). It was decided to use the pUC18 vector as it has the same MCS in the same orientation as the M13mp18 vector and so the Region 3 subcloning strategy described above would be directly applicable to the vector. SSCP analysis had located four mutations in Region 1 and therefore this seemed the most productive region to begin the cloning procedure. Additionally the use of forward and reverse primers would allow the whole of Region 1 to be sequenced using a single clone.

PCR amplification of Region 1 DNA: PCR amplification of wild-type template genomic DNA, Region 1 primers and the conditions which had proved adequate for SSCP analysis failed to yield any visible product when monitored by agarose gel electrophoresis. Use of different sources of or fresh preparation of genomic DNA also yielded no amplified DNA. Likewise use of new Taq polymerase or newly prepared reaction mixtures failed to rectify the situation. Thus it appeared that the primers were at fault. The whole stock of the Region 1 primers had been prepared by E. von Schothorst and stored in water at -20°C for about 3 years. The primers are less stable in water than in the ammonium solution in which they are supplied. Primer set 3, which had only been prepared at quarter volume as required, was working and I concluded that the original Region 1 primers had degraded and they were discarded. While it was possible to reorder the same primers, these would still have required the time consuming and relatively inefficient procedure for cloning the amplified fragments. It was therefore decided to take this opportunity to design a new set of PCR primers adapted specifically for cloning the amplified sequences.

Design of new Region 1 primers: In order to aid the cloning procedure primers were designed that incorporate restriction sites. Use of a particular restriction endonuclease will enable the digestion of the PCR amplified DNA for cloning into that restriction site in the MCS with greater efficiency than the blunt end ligation used previously. Moreover this method requires fewer steps for the preparation of the PCR amplified fragment, thus the loss of DNA would be reduced adding to the efficiency of the cloning procedure. In addition, placing a different restriction site in each of the primers would allow the directional cloning of the PCR product. Sequence data could be obtained for the complete length of one Region 1 strand using a single clone, and subcloning into pUC19 (which has the MCS in the opposite orientation) would yield sequence data for the entire length of the second Region 1 strand. Digestion of the vector with different restriction enzymes will yield a linear molecule with non-compatible ends which will prevent re-ligation of single vector molecules.

A computer aided analysis of the *qutA* sequence -297 to +703 was carried out for those restriction sites present in the MCS of pUC18. An *Eco*RI restriction site (5' GAATC 3') found at -70bp is 10bp 5' of the original Region 1 sense primer. No suitable restriction site could be found at the 3' end of the DNA sequence. However a single base alteration C to G at 537 would create a *Bam*HI site (5' GGATCC 3'). Primers annealing at these positions would amplify a longer fragment than the original Region 1 primers but not so much as to prevent sequencing the whole region from a single clone. The Primers program could not be adjusted to accommodate primers containing mismatched bases and so the primers were designed by inspection, using those same criteria as described in Section 3.1. However attention was also required to several other details in their design. The pair of primers were designed to have roughly the same annealing temperatures (Ta), that of the antisense primer being calculated ignoring the mismatched base and the bases on either side. Care was taken to ensure that no secondary structures could form within the primers and they would not anneal to each other. The restriction sites incorporated into the primers also had to be carefully placed as restriction enzyme digestion at sites close to the ends of molecules is often inefficient due to poor endonuclease binding. This also applied to the choice of restriction sites in the vector MCS as digestion with one endonuclease forms a linear molecule which may be a poor substrate for the second. There are 21bp between the cleavage points of the *Eco*RI and *Bam*HI sites in the MCS, both endonucleases retain 90% catalytic activity when only a single base pair is present between the end of the fragment and the restriction site. Therefore the primers were designed to have at least seven bases between the cleavage point in the restriction site and the end of the fragment.

Finally the primers selected were used in a computer aided search for homology with fungal sequences in the EMBL database. No matches were found which would have annealing temperatures higher than 10^{o}C below the predicted Ta for the primers. The region amplified by

these primers was designated Region 1A. The primers spanned the region -82 to +544: The sense-strand primer -82 to -65 sequence is 5' CCGCCGGAAAA**GAATTC** 3' (bold sequence represents the *Eco*RI site): The antisense-strand primer 544 to 527 sequence is 5' CTTCA**GGATCC**GCTGGGG 3' (bold sequence represents the *Bam*HI site, the underlined **G** is a mismatch incorporated to create the site).

Cloning Region 1A PCR amplified DNA into pUC18: The new primers were prepared. Most of the parameters for their use had previously been established as adequate for amplification from *A.nidulans* genomic DNA (Section 3.1). However the annealing temperature for the new primers had to be determined experimentally. Reactions with annealing temperatures between 50-54^{o}C yielded a large number of fragments of differing sizes due to non-specific amplification from the *A.nidulans* genome. Between 55-56^{o}C the number of bands observed was reduced to four fragments of 1200, 880, 620 & 490bp, the major product being the expected 620bp fragment. Above 56^{o}C no amplified DNA was observed. The non-specific amplification at the lower temperatures was deemed to reflect the mismatched base contained in the antisense primer.

Since the expected 620bp fragment constituted the major PCR product and could be separated from the other bands by gel electrophoresis, it was decided to continue with the cloning procedure. The products of twenty PCR reactions, using the wild-type genomic DNA as template, were combined and the fragments produced separated by electrophoresis using a low melting point agarose gel and two different methods were examined to recover the 620bp fragment DNA.

The 'Gene Clean' method involves cutting the desired band from the gel and melting the agarose at 50^{o}C. The DNA is then bound to glass beads, contaminants are washed away and the DNA eluted from the silica with water. The 'Deathwish' method used previously relies on electrophoresis of the required DNA fragment onto dialysis membrane, which is then removed and washed to elute the DNA, then recovered by ethanol precipitation. The Gene Clean method proved to give a higher yield of DNA (90-95%) than the 'Deathwish' method (maximum 75%). Therefore 'Gene Clean' was adopted as the preferred method for the recovery of DNA from agarose gels. The 620bp fragment of DNA was incubated with *Eco*RI and *Bam*HI in universal buffer at 37^{o}C for 2hr. A phenol extraction was performed, the DNA recovered by ethanol precipitation and resuspended in 20µl water. A 5µl aliquot of the solution and 0.5µg λ *Hin*dIII DNA were used in Gel electrophoresis and the amount of DNA present in the sample estimated by comparison with known amount of marker DNA.

A 1µg sample of vector pUC18 DNA (provided by Dr. B. M. Wilkins laboratory) was incubated with *Eco*RI and *Bam*HI and treated in the same way as described for the PCR amplified

DNA above. The vector DNA sample was recovered from the gel and redissolved in 100µl water. The concentration of the sample was estimated using a 10µl aliquot in gel electrophoresis which revealed approximately 1µg vector DNA in the 100µl sample. The concentration and size of the vector and amplified molecules is known, allowing the calculation of the molarity of the samples. Therefore the ligation reactions were set up to contain a 10:1 molar ratio of insert to vector molecules.

The following control samples were also set up for the ligation reactions: Control 1 contained no DNA; Control 2 contained linear pUC18 DNA digested with *Eco*RI and *Bam*HI; Control 3 contained the original circular pUC18. The ligation products were used to transform strain DH5α. Control 1 was a negative control containing no transforming DNA and thus reveals that the strain and growth media were not contaminated. Control 2 was a positive control for the ligation reaction and contained *Eco*RI/*Bam*HI digested pUC18. The blue colonies observed represent those plasmid molecules that had not been linearised or had re-ligated. Control 3 contained circular pUC18 and was a positive control for successful transformation.

The Experimental plates showed the presence of blue colonies in similar numbers to Control 2 representing the un-cut or re-ligated vector molecules. The experiment yielded a large number of recombinant clones represented by the white colonies.

Analysis of Region 1A clones: Twenty well isolated white colonies were selected and plasmid DNA prepared. A portion of the DNA from each isolate was incubated with *Eco*RI to linearise the cloned DNA. The enzyme was removed by phenol extraction and the DNA recovered by ethanol precipitation. Following restriction digestion with *Bam*HI, gel electrophoresis revealed that all twenty isolates were of a similar size (3.2kb) which is consistent with the expected size of the recombinant clone, 2.6kb vector plus a 600bp insert. Further samples of DNA from sixteen of the twenty isolates were used as template in sequencing reactions using the universal primer and Pharmacia T7 sequencing kit.

Sequencing autoradiographs from each of the samples were analysed (data not shown). It was immediately obvious that none of the sequenced clones contained *qutA* DNA. If the sequencing technique described above was to be used for analysis of the *qutA* mutations some method would have to be found to increase the proportion of *qutA* specific clones and to allow identification of *qutA* Region 1A clones. A number of options presented themselves for example:-

Clones of *qutA* could be identified by DNA hybridisation. It is possible to obtain a *qutA* specific probe which contains the whole of the *qutA* gene. Alternatively the original Region 1 primers used successfully in SSCP analysis could be synthesised anew and used in two ways. The

amplified fragment could be used as a probe to screen potential *qutA* clones; or used to produce template for cloning, however this latter approach would involve a return to the less efficient blunt ended ligation procedure. Finally, a new set of primers could be designed. The problem incurred with primers containing mismatched bases demands that new primers would be completely represent the *qutA* sequence. But, in turn, this would mean that such primers would be subject to the same problem as the original Region 1 primers used in cloning. Whilst considering these options I decided to continue the sequencing exercise and analyse the data to discover if the presumed primer failure can be rationalised.

Part II

Analysis of the putative Region 1 clones

Introduction

PCR amplification of Region 1A had yielded a large number of non-*qutA* sequences. An analysis of these sequences was carried out with three aims in mind: To determine which characteristics of the primers caused the amplification of so many undesired sequences, necessary to develop a strategy for cloning Region 1; to determine, if possible, the origin of the non-*qutA* sequences and any common features causing their amplification by the Region 1A primers; to sequence the clones to provide practical experience in the technique, in the interpretation of sequencing autoradiographs, and a valuable exercise in sequence analysis using programs on the mainframe computer.

4 II. 1 Sequence analysis of Region 1A clones

Sequence data was obtained for sixteen clones. The data themselves are is not shown but a summary of the results is presented in Table 4.1. The sequences were used in computer aided DNA identity searches using the fasta program from gcg packages available on UNIX (Deveraux *et al.*, 1984). Sequences were initially compared against each other, revealing that the clones could be classified into six groups denoted A to F. The sequences A to D were represented only once amongst the clones, whilst there were three representatives of the E sequence and nine clones of the F sequence.

The sequences were compared to the entire EMBL database using the fasta program. In one case the F sequences proved to have originated from *A.nidulans.* Sections showing a high percentage

homology were identified but these sections were only 10-30bp in length. These sections varied and appeared to be the result of chance similarity rather than the identification of functional motifs. The search was then repeated limiting the database search *to A. nidulans* sequences. Some highly homologous 10-30bp sections were observed, again nothing suggesting putative functional motifs was observed. The identities given in Table 4.1 represent the longest regions of homology found. The A to E clones were suspected to represent previously unsequenced regions of the *A.nidulans* genome. Therefore it appears that the majority of the 620bp fragments recovered by PCR amplification were due to non-specific amplification of regions in the *A.nidulans* genome.

The *bimC* sequence was analysed in light of these results. The location of the restriction site within the primers dictates that the primer sequence would not be present in the cloned. The primer in the sense direction of the clones were upstream of the published sequence. The Enos and Morris (1990) *bimC* sequence was visually searched for a sequence that could act as the second priming site. The search identified a site with greater than 50% match (9 out of 17 bases) with the target site for the antisense *Bam*HI primer. The site at 488-505bp (co-ordinates from Enos and Morris, 1990) would produce a fragment 635bp in length, approximately equal to the 626bp fragment predicted for *qutA*.

The sequences of the clones E1-3 (Figure 4.1) and of F1-9 (Figure 4.2) were not identical containing a number of single base differences. The clones had been derived from the same genomic source and the differences presumably reflect errors in the different PCR amplified fragments together with errors made in interpretation of the sequencing autoradiograph. Analysis of the Region 1 PCR amplified *bimC* cloned sequences and published *bimC* sequence (Enos & Morris, 1990) indicates three possible errors in the sequence in the database entry. However, all are in the promoter region and none affect the open reading frame of the gene.

The problem of mis-interpretation of the sequencing autoradiograph would be largely mitigated when analysing *qutA* clones as the reactions would be repeated and the published sequence could be used as a reference when reading the sequence.

4 II.2 Analysis of the Region 1A primers

The Region 1A primers had amplified a large number of non-*qutA* sequences from the *A.nidulans* genome. The antisense primer contained a single mismatched base compared with the *qutA* ORF. A more detailed computer aided homology search of the EMBL database with the antisense primer sequence was performed considering only *A.nidulans* sequences rather than all fungal sequences. The search revealed a few sequences whose predicted annealing temperatures were close to that used in the PCR, indicating that the problem may have arisen at least partially due to non-*qutA* specific hybridisation of the antisense primer.

Table 4.1 **DNA sequence analysis of the putative Region 1 clones**

Clone	Length sequence	Gene/ Organism	Location and extent of homology
A	178	*acvA* *A.nidulans*	6794 to 6884 in ORF (55.3% in 94bp) MacCabe *et al.*, 1991
B	167	*penDE* *A.nidulans*	589 to 704 3' to stop codon (59.2% in 120bp) Montenegro *et al.*, 1990 & Tobin *et al.*, 1990)
C	293	*cprA* *A.niger*	1370 to 147 in ORF (54.5% in 99bp) van den Brink *et al.*, EMBL Z26938
D	204	*benA* *A.nidulans*	-1196 to -1016 5' to ORF (52.2% in 182bp) Oakley *et al.*, 1985
E 1-3	120	*uapA* *A.nidulans*	1772 to stop codon and 38bp 3' (59.2% in 98 bp) Gorfienkiel *et al.*, 1993
F 1-9	358*	*bimC* *A.nidulans*	-142 to 1 in ORF (99% in 142bp) Enos & Morris 1990

The table shows the location and extent of identity are given by co-ordinates related to the translation start codon AUG (+1) of each ORF. * the sequence extends 216 nucleotides 5' of the sequence in the database and therefore represents sequence upstream of that reported.

Figure 4.1 **Alignment of cloned sequences E 1-3**

```
   1                                                          70
E1                              AGG CTGAGAAGGT ACCAGCTCCA GAGCCACCTA TAAGC-GCCA
E2 GATTCTGGTC TCTGCTCCTA GTCT-GGAGG CTGAGAAGGT ACCAGCTCCA GAGCCACCTA TAAGCGGCCA
E3 GATTCTGGTC TCTGCTCCTA GTCTGGGAGG CTGAGAAGGT ACCAGCT

  81                                                       122
E1 GTTATTAACG CTGCTGACAT GACGGAACTA ATACTGGTTC GGGTATGCGT AG
E2 GTTATTAACG CTGCTGAGAT GACGGAACTA ATACTGGTTC GGGTATGCGT AG
```

The figure shows the homologous regions of the E 1-3 sequence derived from analysis of the Region 1 clones in pUC18. Dashes (-) indicate a gap inserted to align the sequences. The bases marked in **bold** are those points at which the sequences differ.

Figure 4.2 **Alignment of cloned sequences F 1-9 and the sequence of the *bimC* gene of *A.nidulans***

```
  1                                                                                                  75
3  C-GTTAAAT CAGCG-CTGC TAGTCAGTCA TGGCAAGATG AGACGGATCA TGTGACGA-C CGAGCCGATG ACTGC
4  C-GTTAAAT CA-CG-CTGC TAGTCAGTCA TGGC-AGATG AGACGGATCA TGTGACGA-C CGAGCCGATG ACTGC
5  CGGTTAAAT CAG-G-CTGC TAGTCAGTCA TGGC-A-ATG AGACGGATCA TGTGACGA-C CGAGCCGATG ACT-C
6  CGGTTAAAT CAGCGTCTGC TAGTCAGTCA TGGC-A-ATG AGACGGATCA TGTGACGAGC CGAGCCGATG ACTGC
7 GCGGTTAAAT CAGCGTCTGC TAGTCAGTCA TGGC-A-ATG AGA--GATCA TGTGACGA-C CGAGCCGATG ACTGC
8 GCGGTTAAAT CAGCG-CTGC TAGTCAGTCA TGGC-AGATG AGACGGATCA TGTGACGA-C CGAGCCGATG ACTGC
    x          ++ x                      x x       +x              +                  +
 76                                                                                  150
3 GATGA CCGCGATCCA GAGTCACATG ATGGTCCAAG CGCTGGGGTT TGTTTATGTT TTCC-TTTCT GGCCAACGGA
4 GATGA CCGCGATCCA GAGTCACATG ATGGTCCAAG CGCTGGGGTT TGTTTATGTT TTCC-TTTCT GGCCAACGGA
5 GATGA CCGCGATCCA GAGTCACATG ATGGTCCAAG CGCTGGGGTT TGTTTATGTT TTCC-TTTCT GGCCAACGGA
6 GATGA CCGCGAT-CA GAGTCACATG ATGGTCCAAG CGCTGGG-TT TGTTTATGTT TTCC-TTTCT GGCCAACGGA
7 GATGA CCGCGAT-CA GAGTCACATG ATGGTCCAAG CGCTGGG-TT TGTTTATGTT TTCC-TTTCT GGCCAACGGA
8 GATGA CCGCGATCCA GAGTCACATG ATGGTCCAAG CGCTGGG-TT TGTTTATGTT TTCC-TTTCT GGCCAACGGA
9              x                              TT TGTTTATGTT TTCC-TTTCT GGCCAAC-GA
B                                           x                  TTTTCT GGCCAACGGA
                                                                x           x
151                                                                          225
1                                           A-GCGAGA CCTCATTGCT TATTTACT-A GTTTG
2                                           A-GCGAGA CCTCATTGC- TATTTACTGA GTTTG
3 ACAACAGCCG ACAACCACCA CA-TTTGCCA CACCGCCCCC TCA-GCGAGA CCTCATTGCT TATTTACTGA GTTTG
4 ACAACAGCCG ACAACCACCA CA-TTTGCCA CACCGCCCCC TCAGGCGAGA CCTCATTGCT TATTTACTGA GTTTG
5 ACAACA-CCG ACAACCACCA CATTTTGCCA CACCGCCCC- TCA-G
6 ACAACAGCCG ACAACCACCA CA-TTTGCCA CACCGCCCC- TCA-GCGAGA CCTCATTGCT TATTTACTGA GTTTG
7 ACAACAGCCG ACAACCACCA CA-TTTGCCA CACCGCCCCC TCA-GCGAGA CCTCATTGCT TATTTACTGA GTTTG
8 ACAACAGCCG ACAACCACCA CATTTTGCCA CACCGCCCCC TCA-GCGAGA CCTCA-TGC- TATTTACTGA GTTTG
9 ACAACAGCCG ACAACCACCA CATTTTGCCA CACCGCCCC- TCA-GCGAGA CCTCATTGCT TA-TTACT-A GTTTG
B ACAACAGCCG ACAACCACCA CATTTTGCCA CACCGCCCC- TCA-GCGAGA CCTCATTGCT TATTTACTGA GTTTG
        +                  x                  x    x            x   x   x     x
226                                                                         290
1 -GTCT TATGCCCA-A TTATTGTTTC TGTACCTAAT TCCTACCGTC TATT-CACAT CAT
2 -GTCT TATGCCCA-A TTATTGTTTC TGTACCTAAT TCCTACCGTC TATT-CACAT CAT
3 -GTCT
4 -GTCT
6 -GTCT TATGCCCA-A TTATTGTTTC TGTACC
7 -GTCT TATGCCCA-A TTATTGTTTC TGTACC
8 -GTCT TATGCCCACA TTATTGTTTC TGTACCTAAT TCCTACCGTC TATT-CACAT CATGGCCGG
9 TGTCT TATGCCCA-A TTATTGTTTC TGTACCTAAT TCCTACCGTC TATT-CACAT CATGGCCGGC CCCC
B -GTCT TATGCCCA-A TTATTGTTTC TGTACCTAAT TCCTACCGTC  ATTTCACAT CATGGCCGGC CCCC
  +             +                                    x   x
```

The figure shows the homologous regions of the F 1-9 sequence derived from analysis of the Region 1 clones in pUC18. Dashes (-) indicate a gap inserted to align the sequences. The sequence in **bold** is the consensus sequence derived by comparison of the clones and the *bimC* sequence (B) from the database. The *bimC* start codon is underlined. The differences in sequence marked x represent probable errors in reading the autoradiograph. The three bold **x**s represent possible errors in the published *bimC* sequence. The differences marked + may reflect variations with in the cloned sequences possibly created during PCR amplification.

The same search pattern was used for the Region 1A sense primer sequence. The search revealed a large number of similar sequences, and it was noted that the same stretch of nucleotides was present in the most of the these sequences. Further analysis revealed that the vast majority of these regions of identity lay in non-coding sequences either in the promoter or downstream of the gene in question. In the latter case the sequence may be in the promoter of an adjacent gene. The homology search was then extended to the *A.niger* and *N.crassa* sequences in the database, and yielded very similar results.

This led to the suspicion that the sense primer had been designed inadvertently covering part of a promoter motif. The non-*qutA* PCR products observed were presumably formed by amplification of other sequences containing the same promoter element.

Comparison of the sequences found upstream of genes concerned allowed the derivation of a 15-16 bp consensus sequence for each of the fungal species tested (Figure 4.3).

The consensus sequence, which shows little homology with any known promoter element, was used in a computer aided search of *A.nidulans* sequences in the EMBL database. It is striking that almost every identity found was outside of the coding sequence of the gene concerned.

The putative promoter element discovered (5' AAGAAAAGG 3') shows little homology with any known promoter element, and was found between ~50 and ~1000bp upstream of a large number of genes. The *A.nidulans* genes containing this motif and its distance to the start codon are shown in Table 4.2. It is possible that the motif represents the binding site for a fungal specific enhancer or transcription factor.

However it must be borne in mind that the majority *A.nidulans* genes whose sequence is reported in the database encode enzymes of general metabolism. Therefore the motif may represent the binding site of a general transcription factor with the apparent bias in the class of genes carrying the motif resulting from bias in the genes sequenced rather than the presence or absence of the motif.

igure 4.3 **The consensus sequence for the putative promoter motif**

he Region 1A sense primer

1)	C	T	T	**A**	**G**	**A**	**A**	**A**	**A**	**G**	**G**	C	C	**G**	C	**C**
2)	g	a/g	A	A	G	A	A	A	A	G	G	N	a/g	g	a	c
3)	g	N	a/g	A	G	A	A	A	A	G	G	a	N	a	a/g	c
4)	A	A	A	A	G	A	A	A	A	G	G	N	N	a	N	N

he table shows (1) the sequence of the Region 1 primer. The bases shown in **bold** are also present in the *A.nidulans* consensus equence. The consensus sequences (2-4) were determined by comparison of the Region 1 primer sequence with the EMBL atabase. The twenty sequences showing the highest degree of homology for each organism were then compared to derive the onsensus sequence. Upper case letters denote 75% or higher occourence of this base within the consensus sequence. Lower ase letters denote 35% or higher occourence of this base within the consensus sequence. N represent a base where no bias in equence was evident.

2) The *A.nidulans* consensus sequence included 16 out of 20 homologies found in the promoter region.
3) The *A.niger* consensus sequence included 11 out of 20 homologies found in the promoter region.
4) The *N.crassa* consensus sequence included 18 out of 20 homologies found in the promoter region.

Table 4.2 **The position of the putative promoter motif in *A.nidulans* genes**

Distance to ORF	Gene	Function (Reference)
-80	*qutA*	QA utilisation (Beri *et al.*, 1987)
-80	*gdhA*	NADP dependant glutamate dehydrogenase (Hawkins *et al.*, 1989)
-100	*uapA*	Uric acid xanthine permease (Gorfinkiel *et al.*, 1993)
-132	*H4*	Histone 4 major subunit (Ehinger *et al.*, 1990)
-149	*aciA*	*amdA* inducible gene (Saleeba *et al.*, 1992)
-180	*gpdA*	Glyceraldehyde-3-phosphate dehydrogenase (Punt *et al.*, 1990)
-203	*yA*	Laccase (Aramayo & Timberlake, 1990)
-230	*pgk*	Phosphoglycerate kinase (Clements & Roberts, 1985)
-243	*nimE*	Cyclin B (O'Connell *et al.*, 1992)
-262,-482	*brlA*	Conidiophore development regulation (Prade & Timberlake, 1993)
-600	*facA*	Acetyl Co Enzyme A synthase (Katz & Hynes, 1989)
-600	*wetA*	Laccase (Marshall & Timberlake, 1990)
-670	*lamB*	Lactamase utilisation (Katz & Hynes, 1989)
-886	*crnA*	Nitrate transporter (Unkles *et al.*, 1991)
-991	*tubC*	β–tubulin (May & Morris, 1986)

The table shows some of the genes whose promoter contains a sequence homologous to the putative promoter sequence (AAGAAAAGG). The homologies and their location were identified using the 'Fasta' homology search program on the UNIX computer system. A separate exercise involving the retrieval of identified sequences was required to relate the homology to the ORF of the identified sequence. The adenine of the ATG start codon is designated +1.

Discussion

The amplification of non-*qutA* sequences by Region 1A primers: The amplification of so many non-*qutA* sequences by PCR using the Region 1A primers was clearly the result of the non-specific binding of the sense primer. The primer sequence was related to a sequence found upstream of a number of other genes, and is therefore suspected to be a previously unidentified promoter motif. The location of the motif in the promoter of *lamB* does not overlap any of the identified *getA* and *amdR* or *areA* binding sites (Richardson *et al.*, 1992). The location of the promoter motif in the *qutA* promoter suggests that the motif may be the binding site for a carbon catabolite responsive wide domain regulatory protein. Although it seems more probable that the motif is the binding site for a more general transcription factor. Planned 5' transcript mapping for the *qutA* gene may reveal if this putative promoter has any functional significance in *qut* regulation. The reason why previous identity searches for the proposed primer sequences failed to uncover the homologies described above is now apparent. All fungal sequences in the database were tested and these include the vast amount of sequence data from *S.cerevisiae*. This broad search discovered genomic sequences with greater identity to the proposed primers in organisms other than *A.nidulans*, but in turn these revealed no consistent pattern in the bases with identity.

Implications for amplification of *qutA* Region 1: The analysis of the Region 1A clones made it clear that the cloning procedure for analysis of *qutA* mutations required improvement. Sequence comparison of the E and F clones with each other and the sequence retrieved from the EMBL database had detected a number of single base differences (Figures 4.1 and 4.2). Eighteen of the 24 apparent differences could be explained as errors in reading the sequencing autoradiograph, a problem that would be reduced when reading *qutA* sequences since the published sequence is available as a reference. However, the 8 probable PCR errors in a total of 1,696bp sequenced (1 every 212bp) together with discussion with colleagues engaged in similar techniques, suggests that between 5 and 10 clones of each mutant strain would have to be analysed to ensure any sequence alteration identified is real and not a PCR artefact. Such numbers would require an efficient cloning procedure.

Chapter Five

DNA sequence analysis of *qutA* Region 1

The work in Chapter 4 describes a method by which it is possible to clone PCR amplified DNA for sequence analysis. However this approach had not yielded any *qutA* sequence data, due to an unforeseen problem with the sense primer amplifying other *A.nidulans* promoter regions. It was decided to design a further set of Region 1 primers placed outside the Region 1A primers.

Recent developments in sequence analysis based upon PCR suggests that amplified dsDNA can be treated as template in sequencing reactions in the same way as plasmid DNA. A potential disadvantage is that the PCR primers may prove ineffective as sequencing primers, as I found in attempting the direct sequencing of PCR products described in Section 4.1. However the new external primers could be used to prepare the *qutA* specific PCR amplified dsDNA and the Region 1A primers would then be available as internal primers for sequencing. This strategy would counteract the inherent risk that a single cloned sequence used as a template might carrying base errors. Potential PCR errors would still be present, but the results from the sequencing reactions would be based upon a population of PCR amplified fragments and not a single fragment. An error occurring early in PCR still might be highly represented in the final population of molecules. However this situation would be detected in the sequence data.

5.1 Direct sequence analysis of PCR amplified dsDNA

The new external primers, designated Primer set 1B, were designed for direct sequencing of PCR amplified dsDNA template. If the method fails to yield sequence data, the primers would be available to amplify *qutA* specific ds DNA for cloning in to a vector.

Design and use of the *qutA* Region 1B PCR primers: The new primers do not require the incorporation of mismatched bases, and so the 'Primers' program could be used in their design. The size of the amplified fragment was not critical except that the primer positions lie outside of the Region 1A primer sites. The reference sequence -297 to +703 and the criteria previously adopted were examined (Section 3.1). Because the primers sequences are longer a higher annealing temperature can be used than with the Region 1A primers, which should increase the specificity of primer binding. The sense primer was placed at the furthest 5' position available (-294 to -272) in the *qutA* sequence in the hope of avoiding all core promoter elements. Additionally, it may allow

the identification of one or more of the mutations predicted to fall in the promoter region. The sequences of Region 1B primer and their relationship to the *qutA* gene is given below. Primer set 1B spanned the region -294 to +599: Sense-strand primer -294 to -272 sequence 5' AATTCCCCAACTCCTCTTCT 3': Antisense-strand primer 599 to 578 sequence 5' GCGATAGAATCTGGAAGCTGCC 3'.

Conditions of the PCR: Wild-type genomic DNA was used as template when determining the conditions for the amplification of Region 1B by PCR. All the conditions remained as previously described (Chapter 3.1) except for the following:- Region 1B is 894bp long and the time for DNA synthesis was increased to 1.5min. The annealing temperature was varied between 50-67^{o}C, and at 67^{o}C the PCR resulted in the amplification of a single band of the expected size.

Preparing Region 1 template for sequencing reactions: The products of ten parallel PCR reactions amplifying wild-type (*R153*) genomic DNA using Primer set 1B were combined. Excess primer and unincorporated nucleotides were removed by gel electrophoresis and the amplified DNA recovered by the 'Gene Clean' method (Section 2.6.1). The quantity of amplified DNA was determined by electrophoresis of an aliquot of the DNA recovered and comparison with known quantities of λ *Hind*III DNA. A yield of 1.7mg of dsDNA template was obtained.

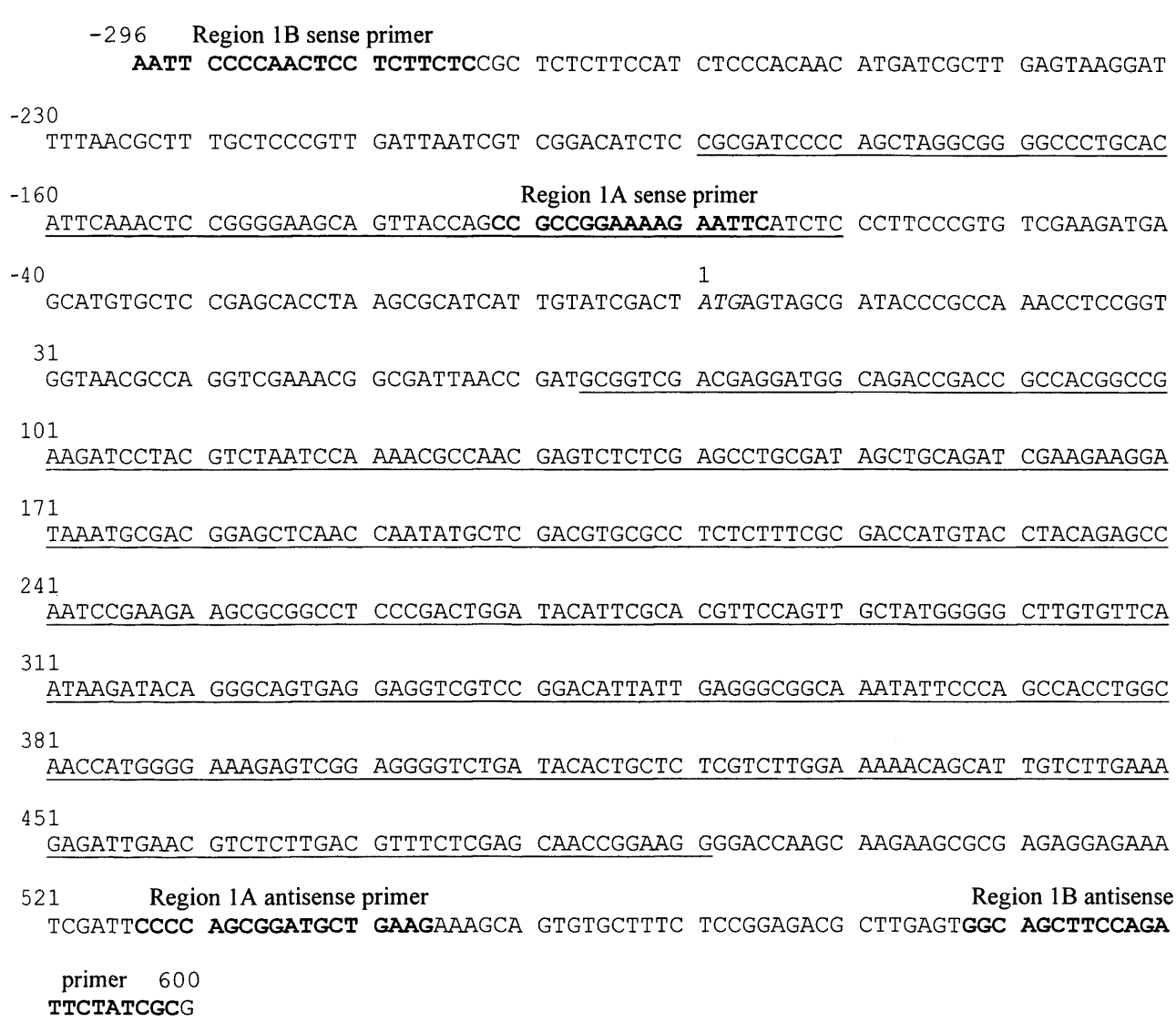
The template DNA was diluted to contain 100ng in each sequencing reaction. The Region 1A and 1B pairs of primers were diluted to 0.5pM, each was used separately as a sequencing primer in a series of four reactions with the Region 1B dsDNA template and the Pharmacia T7 sequencing kit. Sequencing reactions were carried out as described in Section 2.9.3 and the sequencing gel prepared as in Section 2.9.4.

Bands in each of the sequence ladders were rather faint suggesting that more template should be used in subsequent reactions (Data not shown). However it was possible to obtain some sequence data in each case and this confirmed that the template was *qutA* Region 1. The Region 1 sequence is shown in Figure 5.1. with the sequence obtained with each primer. The sequence data matched the published *qutA* sequence perfectly and no PCR errors were apparent. Surprisingly, the outer Region 1B primers gave clearer sequence data than the inner Region 1A primers, showing that these PCR primers could also be used as sequencing primers despite their proximity to the ends of the amplified fragment.

Since I had at last found an effective sequencing strategy I decided not to pursue the cloning procedure in favour of the direct sequencing method described here.

The sequence reactions described above had indicated that more than 100ng template dsDNA was required for each sequencing reaction. Pooling five PCR reactions would provide approximately 1mg of template, which would allow four sets of sequencing reactions containing

Figure 5.1 **Direct sequencing of *qutA* Region PCR amplified dsDNA**



The figure shows the sense strand sequence of Region 1B. The primer sequences are shown in **bold**. The sequence underlined represents that obtained during sequencing reactions. The extent of sequence obtained from each primer is given below. Region 1B sense primer (co-ordinates -294 to -272) extent of sequencing -150 to -60. Region 1A sense primer (co-ordinates -81 to -64) extent of sequencing +64 to +275. Region 1B sense primer (co-ordinates +543 to +526) extent of sequencing +263 to +490. Region 1B sense primer (co-ordinates +599 to +578) extent of sequencing +244 to +491.

250ng template in each reaction. This scale should be sufficient to produce clear sequence data. When a mutation is discovered the procedure will be repeated fully, with the amplification of new template dsDNA to confirm the result obtained and to provide an additional control against mis-identification of PCR errors.

5.2 Sequencing mutations mapped to Region 1B by SSCP analysis

The first targets for sequence analysis were the mutations located in Region 1 by SSCP analysis (*qutA423*, *422 400* and *552*). Of these *qutA422* mapped to the smallest segment and *qutA423* mapped to an overlapping restriction fragment. Sequencing template was prepared by PCR amplification of Region 1 using the 1B primers from genomic DNA of these strains. The sequencing reactions were carried out using the both Region 1A and 1B sense primers. Sequence data obtained using the Region 1A sense primer identified mutations in strains *qutA423* and *422*. The *qutA422* mutation an additional A at bp 44 is shown adjacent to the wild-type sequence at the same region in Figure 5.2(A). The results are summarised in Table 5.1.

The mutations *qutA400* and *552* also fell in overlapping restriction fragments and were sequenced at the same time. Template was prepared and reactions carried out as described above. In this case the Region 1B antisense primer was used as the sequencing primer as it had provided clearer sequence data than the Region 1A antisense primer in analysis of wild-type sequence. Sequence data identified the mutations in strains *qutA400* and *552*. The *qutA400* mutation, a deleted T residue, and the *qutA552* mutation, a C to T substitution, are shown adjacent to the wild-type sequences for the same regions in Figure 5.2B and C respectively. The results are summarised in Table 5.1.

Analysis of the sequencing results: The *qutA423* and *422* mutations proved identical, each resulting from the addition of an adenine residue (A) at bp +44. The mutation lies within the fragments identified by SSCP analysis for both strains. Each mutation had been mapped to the same *Hpa*II restriction fragment, however the *Hae*III digest which had refined the location of *qutA422* to bp +27 to +98 was not used in the analysis of *qutA423*. Thus mapping the mutations to different overlapping restriction fragments by SSCP analysis was not inconsistent with the identity of the two mutations. The result was puzzling in that the probability of two separately isolated mutant strains carrying the same mutation is very small. Mutants were isolated by replica plating from dishes bearing up to 75 colonies (Grant *et al.*, 1988). It is understandable that the same colony might be mistakenly isolated twice and the consecutive numbering of the strains reinforces this interpretation.

Figure 5.2 **Sequencing autoradiographs of mutations identified in Region 1 by SSCP**

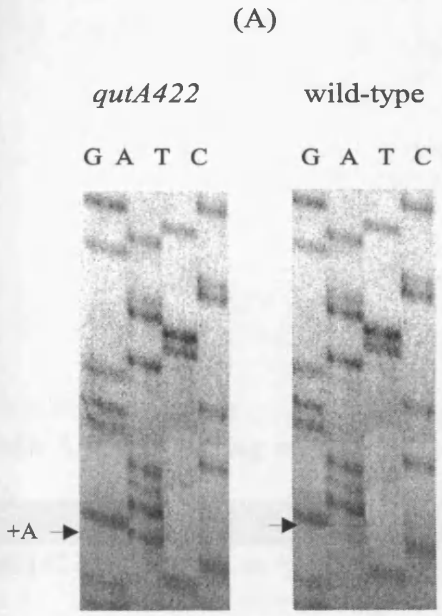


Figure (A) shows the addition of an A in the *qutA422* sequence compared to the wild-type *qutA* gene.

wild-type 40 to 65 GGTC GAAACGGCGATTAACCGATGC
quta422 40 to 65 GGTCAGAAACGGCGATTAACCGATGC

The mutation was identified in Region 1 using the sense primer. The mutation is therefore an additional A at bp 44.

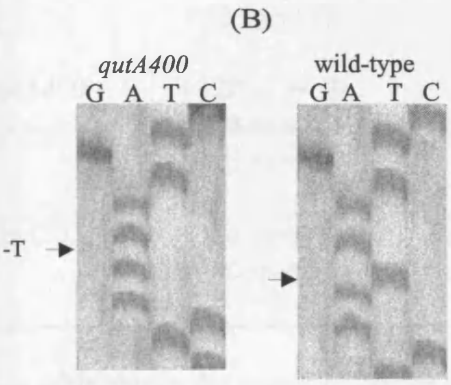


Figure (B) shows the deletion of a T in the *qutA400* sequence compared to the wild-type *qutA* gene.

wild-type bp 351 to 340 TCAATAATGTC
quta400 bp 351 to 340 TCAA AATGTC

The mutation was identified in Region 1 using the antisense primer. The mutation is equivalent to a deleted A at bp 348.

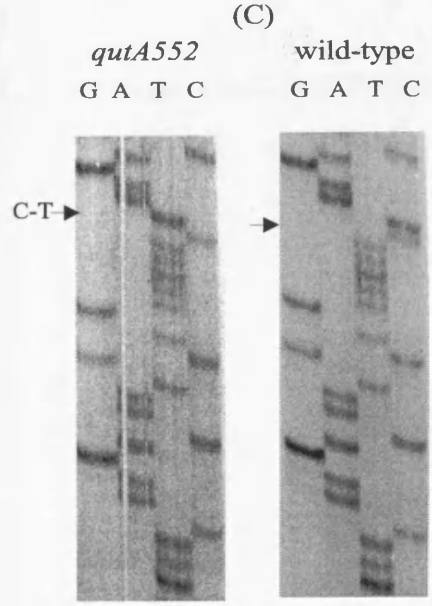


Figure (C) shows the substitution of a C to T in the *qutA552* sequence compared to the wild-type *qutA* gene.

wild-type bp 449 to 421 TTTCAAGACAATGCTGTTTTTCCAAGAC
quta552 bp 449 to 421 TTTCAAGACAATGCTGTTTTTCTAAGAC

The mutation was identified in Region 1 using the antisense primer. The mutation is equivalent to a G to A substitution at bp 428.

Table 5.1 **Sequencing mutations mapped to Region 1 by SSCP analysis of the *qutA* gene**

Strain	Extent of sequencing	Mutation identified	Location of mutation by SSCP
qutA422	+13 to +123	Additional A at +44	+27 to +98
	+25 to +100	Additional A at +44	
qutA423	+12 to +117	Additional A at +44	+27 to +209
	+25 to +100	Additional A at +44	
	+53 to +179	No mutation found	
qutA400	+320 to +450	Deleted A at +348	+340 to +484
	+324 to +465	Deleted A at +348	
	+415 to +557	No mutation identified	
qutA552	+324 to +465	Substituted A to G at +428	+258 to +484
	+360 to +500	Substituted A to G at +428	

The table shows the extent of sequence obtained from each mutant strain, the mutation identified and the location of the mutation as identified by DNA sequence and SSCP analysis. The co-ordinates and base alteration refer to the sense strand of the *qutA* ORF.

The data also displays the limitations of genetic analysis. Both *qutA422* and *qutA423* had been crossed to the marker strain *qutA552* resulting in respectively 63 and 70 recombinants per million spores plated. Therefore a difference in RF between two test strains of less than 10 recombinants in 10^6 spores must be considered insufficient to separate mutations.

The *qutA400* and *552* mutations had been mapped to different but overlapping restriction fragments by SSCP analysis. The sequence data placed each at separate points within this region. The alignment of the genetic and physical maps had placed *qutA400* 5' to *qutA552* (Figure 3.9) and the sequencing results confirms this result; *qutA400* at +348, *qutA552* at +428. At this point in the work the alignment and co-linearity of the genetic and physical maps appears satisfactory.

5.3 Mutations predicted to map to Region 1 by the genetic map

Sequence data for Region 1 can be obtained using both the Region 1A and 1B primers, extending over a range approximately -250 to +550 in the *qutA* gene. The SSCP analysis described in Chapter 3 and DNA sequence analysis has located four mutations (*qutA422*, *423*, *400* and *552*) within this region and the alignment of the genetic and physical maps predicts that a further seven mutations lie within these sequence co-ordinates (Table 3.4). Two mutations *qutA380* and *402* were predicted to lie in the promoter region of *qutA* but further upstream than the Region 1B sense primer. I had no plan to sequence further upstream of *qutA* than this, however, the two strains were included for analysis with the most 5' mapping primer. Genomic DNA was prepared from each of the strains for use as template in Region 1B PCR reactions.

Mutations predicted to map to the 5' end of Region 1: Of the three strains (*qutA380, 402* and *404*) predicted to lie in the promoter only *qutA404* is predicted to lie within Region 1B. However PCR amplified DNA from each strain was used in sequencing reactions with the Region 1B sense primer. Among the three strains tested only the *qutA402* mutation (-237) was detected and is shown adjacent to the wild-type sequence at the same region in Figure 5.3 (A) The mutation lies as predicted in the promoter region of the *qutA* gene (Table 5.2).

Analysis of the sequence data: The *qutA380* mutation was not found, as expected, and is assumed to lie 5' (before -265) to Region 1. The *qutA402* mutation was identified as the addition of a T at -237. The genetic map places the *qutA404* mutation 3' of *qutA402* and therefore it was expected to lie within Region 1B, but no mutation was found for the strain *qutA404*. It could be that the mutation was simply missed in the same way as discussed for *qutA341*. However in this case it seems more likely that the genetic map had misplaced the *qutA404* mutation. The genetic map is

Figure 5.3 **Sequencing autoradiographs of mutations identified in Region 1**

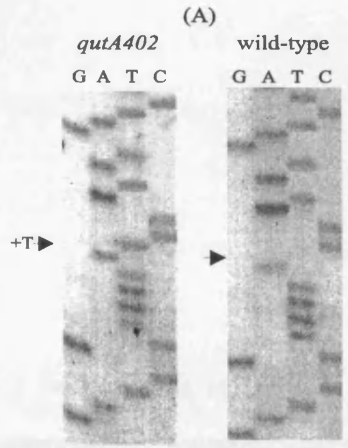


Figure (A) shows the addition of a T in the *qutA402* sequence compared to the wild-type *qutA* gene.

wild-type bp -248 to -227 GATCGCTTTTA CCATATGATC
qutA402 bp -248 to -227 GATCGCTTTTATCCATATGATC

The mutation was identified using the Region 1B sense strand primer. The mutation is an additional T residue at bp -237.

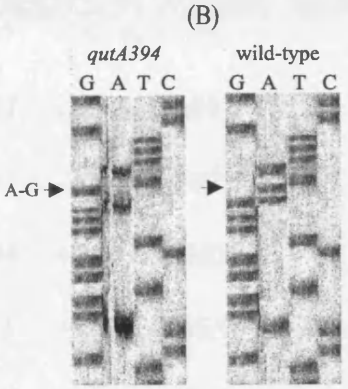


Figure (B) shows the substitution of an A to G in the *qutA394* sequence compared to the wild-type *qutA* gene.

wild-type bp 382 to 355 TTGCCAGGTGGCTGGGAATATTTGCCG
qutA394 bp 382 to 355 TTGCCAGGTGGCTGGGAGTATTTGCCG

The mutation was identified using the Region 1B antisense primer. The mutation is equivalent to a T to C substitution in the sense strand at bp 364.

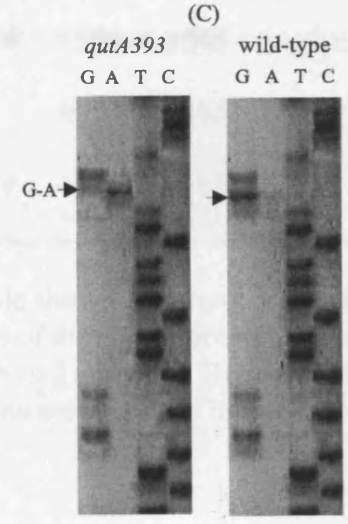


Figure (C) shows the substitution of a G to A in the *qutA393* sequence compared to the wild-type *qutA* gene.

wild-type bp 512 to 490 TCTTCGCGCTTCTTTCTTGGTCCT
qutA394 bp 512 to 490 TCTTCGCGCTTCTTTCTTAGTCCT

The mutation was identified using the Region 1B antisense primer. The mutation is equivalent to a C to A substitution in the sense strand at bp 495.

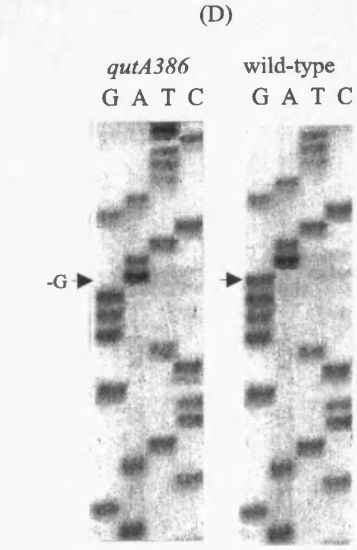


Figure (D) shows the deletion of a G in the *qutA393* sequence compared to the wild-type *qutA* gene.

wild-type bp 537 to 517 AGCATCCGCTGGGGAACTGATTT
qutA394 bp 537 to 515 AGGATCCGCTGGG AACTGATTTCT

The mutation was identified using the Region 1B antisense primer. The mutation is equivalent to a deleted C in the sense strand at bp 529.

Table 5.2 **Sequencing mutations predicted to map to Region 1**

Strain	Extent of sequencing	Mutation identified	Predicted location
qutA380	-265 to +189	(5' to -265)	-1440
qutA402	**-265 to +189**	**Additional T at -237**	**-386**
qutA404	-265 to +189	(5' to -265)	-219
qutA394	**+350 to +563**	**Substituted T to C at +364**	**+322**
qutA341	+350 to +565	(+428 to +496)	+361
qutA393	**+489 to +556**	**Substituted C to T at +495**	**+379**
qutA386	**+350 to +565**	**Deleted C at +530**	**+410**
qutA4	+350 to +565	(3' to 565)	+436
qutA384	+350 to +565	(3' to 565)	+458

The table shows the extent of sequence obtained from each mutant strain, the mutation identified and the location of the mutation as identified by DNA sequence analysis. The predicted location of mutations is taken from Table 3.4. The co-ordinates and base alteration refer to the sense strand of the *qutA* ORF. Where no mutation has been identified a probable location is noted in parenthesis.

inconsistent in the orientation of the *qutA404* and *402* mutations; the crosses to *qutA552* had placed *qutA402* upstream of *qutA404* (this data was used in compilation of the Genetic map, Figure 1.5). However the crosses to *qutA506* give the reverse order and this appears to be in agreement with the sequencing results. Therefore both the *qutA380* and *qutA404* mutations probably lie 5' to Region 1, that is before -265 with respect to the translation start codon.

Mutations in the 3' end of Region 1 of the *qutA* gene: Most of the mutations predicted to lie in Region 1 by the alignment of the genetic and physical maps are placed in the 3' half of Region 1, and therefore within sequencing range of the Region 1B antisense primer. Sequencing template was prepared by PCR from genomic DNA of each of the mutant strains *qutA394*, *341*, *393*, *386*, *4* and *384*. Sequencing reaction were carried out using the Region 1B antisense primer. Mutations were identified in three of the six strains *qutA394, 393* and *386*, and these results are summarised in Table 5.2.

Analysis of the sequence data: The genetic map predicts that the *qutA394* mutation (+364) lies between *qutA400* (+348)and *qutA552* (+428) and this proves to be the case. The mutation was identified as the deletion of an A residue using the antisense primer (Figure 5.3B). The location of each of the mutations *qutA393* and *386* were also consistent with the genetic map order. The *qutA393* and *386* mutations are shown adjacent to the wild-type sequence at the same region in Figure 5.3 (C) and (D) respectively. No sequence changes were detected for the mutant strains *qutA4* and *384*, and these mutations presumably lie 3' to Region 1B. However the mutation *qutA341* expected to fall in this region was not found.

It is possible that the genetic map has misplaced *qutA341*. However this seems unlikely as the genetic data from the crosses *qutA341* with marker strains *qutA552*, *506* and *636* were consistent in their location of *qutA341*. Alternately the mutation is present in the region sequenced but had not been identified. If the mutation was the addition or deletion of one base in a series of similar bases, then it may not have been detected if it fell in apart of the sequencing gel less favourable for analysis.

Discussion

The method for sequencing described above proved both rapid and reliable. The method has several advantages over a cloning procedure. It required less processing of the amplified dsDNA before sequencing and was therefore more efficient in the use of PCR amplified DNA. More importantly it produced *qutA* sequence data and no obvious PCR errors have been identified. Each of the eight mutations identified was confirmed by repeating both the PCR amplification of the dsDNA template and a second set of sequencing reactions. In each case the mutation was clearly identifiable and the data is therefore of high quality and reliable.

One mutation, *qutA341,* expected to lie within Region 1 was not located by DNA sequencing. It is felt that this was probably due to difficulties in observation difficulties rather than its location elsewhere. A mutation involving a single base addition or deletion in a series of similar bases may easily be overlooked unless in the most favourable part of the sequencing gel.

Eight mutations had been identified in Region 1B (*qutA402*, *422*, *423*, *400*, *394*, *552*, *393* and *386*). It is satisfying that the genetic and physical maps prove co-linear.

Two mutations *qutA404* and *402* were placed physically in a reversed order compared to the original genetic map. However the genetic data itself was inconsistent in that *qutA404* and *402* had been placed in the opposite order in crosses. Sequence analysis has identified the *qutA402* mutation but not *qutA404*, which presumably lies 5' of Region 1.

Two mutant strains *qutA422* and *423* proved to have identical mutations. Two other examples of consecutive numbered mutants that map genetically close to each other are evident from the genetic map. The *qutA384* and *385* that map near each other but are definitely not the same strain as they have different phenotypes. The *qutA389* and *390* mutations do share the same phenotype and map genetically very close together, and these may also be separate isolates of the same strain.

The sequencing result of strains *qutA422* and *423* gave further insight into the limitations of the genetic analysis. The ordering of mutant and marker strains was definitive. However, the RF between strains was not sufficient to order close mapping mutations, and the predicted order of mutant strains using genetic map distances of less than 10 recombinants in 10^6 spores plated could not be considered sufficiently reliable to determine their order.

The physical location of the five mutations described above allowed a re-calculation of the bp/RF ratio based upon actual rather than estimated physical positions (Table 5.3). The figure obtained 4.7 bp: 1 recombinant in 10^6 spores plated is similar to that used previously (4.3 bp/RF) in the prediction of physical locations, and was not felt to warrant a major revision of the predicted position of mutations. The final calculation of bp:RF ratio is shown in Chapter 7.

Among five mutant strains *qutA422, 423, 400, 394,* and *552* tested by SSCP analysis, four were mapped to restriction fragments, while that in *qutA394* mutation was not detected. Sequencing has shown that all five mutations lie within Region 1. Comparison of the characteristics of mutations located by SSCP analysis with those not detected may prove of value in judging the limitations of the technique. However since further relevant data also arises from sequencing mutations in Region 3, this matter will be discussed in the final chapter when all of the data will be available.

Mutant *QUTA* activator proteins: The changes in protein sequence predicted by DNA sequence analysis of Region 1 mutations were determined both by inspection and by a DNA sequence translation program on the Genetic Computer Group software package on the university mainframe IRIX system. The results are shown in Table 5.4. The predicted proteins formed by the identified *qutA* mutations will be discussed in detail in Chapter 8, when all the available data can be considered together. Three major observations result from sequencing *qutA* mutations in Region 1 of the activator gene.

The genetic and physical maps are co-linear, this result is expected, but is particularly important for the selection of mutant strains for sequence analysis of further regions. **The alignment of the genetic and physical maps still places no mutations in the 3' portion of the *qutA* gene,** therefore no mutant strains have been found with changes in the carboxyl terminal half of the activator protein. **No mutations mapped within the sequence encoding the putative DNA binding domain of the *QUTA* protein.** The last two features raise major questions about the functional significance and position of putative motifs in the *QUTA* activator protein.

Table 5.3 **Calculation of the bp/RF ratio using actual physical distances**

Strain	RF	bp	bp:RF	Strain	RF	bp	bp:RF
402* to *552	172	665	3.9	***402* to *506***	242	1368	5.7
422* to *552	67	384	6.1	***422* to *506***	232	1087	4.7
400* to *552	19	80	4.2	***400* to *506***	174	783	4.5
394* to *552	3	64	21	***394* to *506***	245	767	3.1
393* to *552	10	68	6.8	***393* to *506***	193	635	3.3
386* to *552	17	103	6	***386* to *506***	151	601	4
Mean = 4.7							

The RF values are taken from Appendix 2. The *qutA552* and *qutA506* marker strains could not be crossed to each other and so no RF value is available. The physical distances used are actual distances between mutations accept for the *qutA506* mutation which is estimated to lie in the centre of the restriction fragment to which it was mapped by SSCP analysis (Chapter 3).

Table 5.4 **Translation of mutant *qutA* sequences identified in Region 1**

Strain and phenotype	Mutation	Alteration to protein	Protein length
qutA402 SD	Additional T at -237	Frameshift bringing a 5' OFR into frame with the QUTA gene	108 aa added 5'to 825 wt
qutA422 SD	Additional A at 44	Frameshift incorporating 23 non-QUTA aa before termination	23 aa added 3'to 15 wt = 38
qutA400 SD	Deleted A at 348	Frameshift incorporating 1 non-QUTA aa before termination	1 aa added 3'to 116 wt = 117
qutA394 SD	Substituted T to C at 364	Single amino acid substitution I to T at residue 122	825
qutA552 R	Substituted G to A at 428	Nonsense mutation	142
qutA393 R	Substituted G to T at 495	Nonsense mutation	164
qutA386 SD	Deleted C at 529	Frameshift incorporating 30 non-QUTA aa before termination	30 aa added 3'to 177 wt = 207

The position of the mutations refer to the *qutA* ORF. The amino acid (aa) alterations refer to the amino acid sequence of the wild-type (wt) QUTA protein (Appendix 1). The recessive phenotype is represented by R and the semidominant phenotype by SD.

Chapter Six

DNA sequence analysis of Region 2 of the *qutA* gene

Chapter 5 described a successful approach to sequencing mutations in the *qutA* gene. Sequence data from Region 1 and the predicted location of mutations (Table 3.4) indicates that mutations lie between Regions 1 and 3. This chapter describes the design and use of primers for the PCR amplification and sequencing analysis of this intervening region, designated Region 2.

Design of the Region 2 primers: The Region 2 primers were designed using the 'Primers' program with the reference sequence +302 to +1102 and the criteria discussed previously (Section 5.1). The primers were chosen to amplify a region of *qutA* short enough to be allow sequence analysis of the intervening region by using the PCR primers. However Region 2 had to be long enough to allow the design of internal primers if the PCR primers could not be used as sequencing primers for the region. As some overlap was necessary between Regions 1, 2 and 3 the primers were designed such that mutations mapping to the 3' end of Region 1 and 5' end of Region 3 would also be expected to lie within Region 2. This would provide sequence data for the complementary strand of these mutations by using separate amplified fragments with different primers. Together, this approach will provide an additional gauge for the reliability of the sequencing technique. The Region 2 primers spanned the sequence 404 to 1078, their positions in the *qutA* gene are given below: Sense-strand primer sequence (404 to 423) 5' GTCTGATACACTGCTCTCG 3': the antisense-strand primer sequence (1078 to 1057) is
5' TGTTGGGATAAGACTCCGTCG 3'.

PCR amplification of Region 2 sequencing template: The new primers were designed for use as both PCR and sequencing primers. The conditions of the PCR amplification of Region 2 are as described in Section 3.3 except for the alterations described below. Region 2 is 647bp long and 1min DNA synthesis was considered ample. The annealing temperature was varied between 50-58^{o}C, and at 58^{o}C the PCR yielded a single amplified fragment of the expected size. The amplified DNA from five parallel PCR reactions of each of the genomic DNA source were combined to provide the template for sequencing reactions and stored at 4^{o}C until sequencing reactions were carried out.

6.1 Choice of mutant strains for analysis in Region 2 of the *qutA* gene

Sequence analysis of Region 1 had identified base changes in the mutant strains *qutA552, 341, 393* and *386,* which also lie within the Region 2 amplified sequence. The alignment of the genetic and physical maps (Table 3.4) predicts that a further six mutations (*qutA4, 384, 399, 385, 633* and *379*) lie within Region 2 (+423 to +1057). The mutation *qutA552* lies too close to the primer binding site for sequence analysis using Region 2 template. The remaining nine strains were chosen for sequence analysis of Region 2. Genomic DNA was prepared from each strain for use as template in PCR amplification of Region 2.

Sequence analysis of mutations mapping to the 5' end of Region 2

The first targets for sequence analysis of Region 2 were the three strains previously tested in analysis of Region 1. That is strains *qutA393* and *386* in which the location of the mutation is known and the strain *qutA341* for which no mutation had been identified. The two strains *qutA4* and *384* were expected to contain mutations a short distance 3' of bp 565, which is within sequencing range of the Region 2 sense primer. Sequencing reactions were carried out with the Region 2 amplified DNA template from each of these five strains using the Region 2 sense primer. The location of the previously mapped mutations was confirmed and one new mutation, *qutA4* ,was identified (Figure 6.1A). The results of the sequencing are summarised in Table 6.1.

Analysis of the sequence data: The Region 2 sense PCR primer yields sequence data for Region 2. Sequence analysis of *qutA341* again failed to identify the mutation, this could be simply because the mutation was missed again or that it lies 5' to bp 464. It is satisfying that both the *qutA393* and the *386* mutations were each found by sequencing the complementary (sense) strand to that sequenced in the work using the Region 1 primers. In both cases the mutational changes were identical to those previously found (*qutA393* a G to T substitution at 496; *qutA386* a deleted C at 529). Additionally the *qutA4* mutation was identified (deletion of A at 613) and as predicted lay just 3' of Region 1. The *qutA384* mutation was not found and probably lies 3' of bp 694.

Sequence analysis of mutations mapping to the 3' end of Region 2

The mutations in strains *qutA399, 385, 633* and *379* are predicted to lie in the 3' half of Region 2 and *qutA384* was predicted to lie 3' of bp +694. Each of these is hoped to lie within sequencing range of the antisense primer. Sequencing reactions were carried out using the Region 2 antisense primer and mutations were identified for three of the strains tested *qutA385* (Figure 6.1 B), *633* and *379*. The results are summarised in Table 6.1.

Figure 6.1 **Sequencing autoradiographs of mutations identified in Region 2**

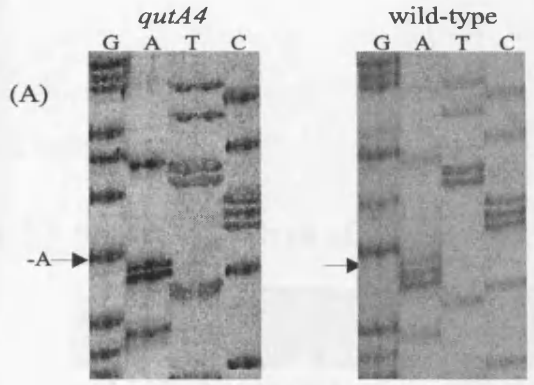


Figure (A) shows the deletion of an A in the *qutA4* sequence compared to the wild-type *qutA* gene.

wild-type bp 602 to 629 TGCGAGTCAAAGCCCGTTAGCGTCTGGG
qutA4 bp 602 to 629 TGCGAGTC AAGCCCGTTAGCGTCTGGG

The mutation was identified in Region 2 using the sense primer. The mutation is therefore a deleted A at bp 611.

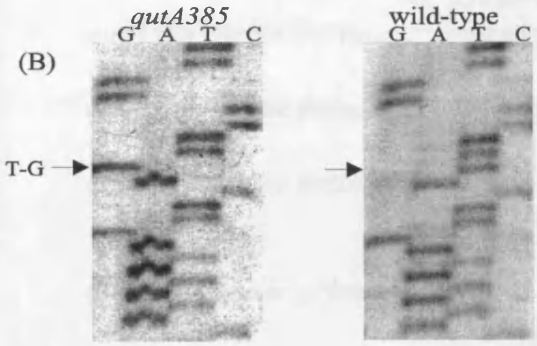


Figure (B) shows the substitution of a T to G in the *qutA385* sequence compared to the wild-type *qutA* gene.

wild-type bp 824 to 803 CATATATAGCTTCATTTCCGGTT
qutA4 bp 824 to 803 CATATATAGCTTCAGTTCCGGTT

The mutation was identified in Region 2 using the antisense primer. The mutation is therefore equivalent to an A to C subsittution at bp 812.

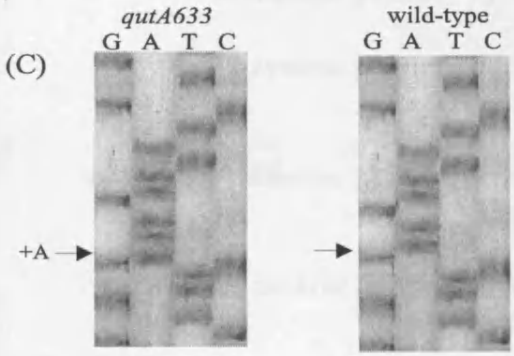


Figure (C) shows the addition of an A in the *qutA633* sequence compared to the wild-type *qutA* gene.

wild-type bp 883 to 863 GCTGTTCGAA GAATATCGTGT
qutA633 bp 883 to 863 GCTGTTCGAAAGAATATCGTGT

The mutation was identified in Region 2 using the antisense primer. The mutation is therefore equivalent to an additional A at bp 876.

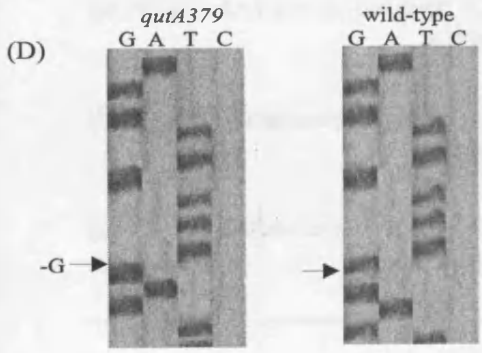


Figure (D) shows the deletion of a G in the *qutA379* sequence compared to the wild-type *qutA* gene.

wild-type bp 987 to 973 TGAGGTTTGTTGGA
qutA633 bp 987 to 973 TGAGGTTTGTTGGA

The mutation was identified in Region 2 using the antisense primer. The mutation is therefore equivalent to an additional A at bp 984.

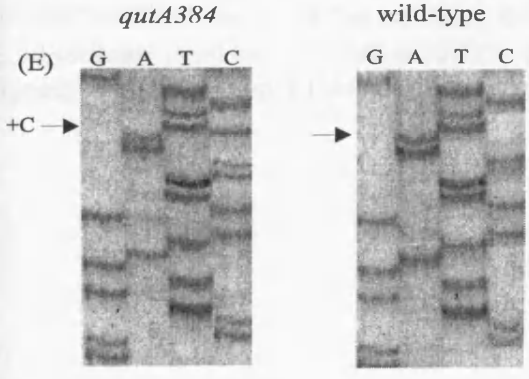


Figure (E) shows the addition of a C in the *qutA384* sequence compared to the wild-type *qutA* gene.

wild-type bp 748 to 771 GGCCTGTGATCGCTTCCAA TTCTG
qutA633 bp 748 to 771 GGCCTGTGATCGCTTCCAACTTCTG

The mutation was identified in Region 2 using the mid sense primer. The mutation is therefore equivalent to an additional A at bp 766.

Table 6.1 **Sequence analysis of Region 2**

Strain	Extent of sequencing		Mutation identified	Predicted /identifed location
qutA341	**Sense primer**	464 to 590 & 556 to 713	none	3' to 464
qutA393	**Sense primer**	463 to 667	Substituted G to T at 495	Substituted G to T at 495
qutA386	**Sense primer**	456 to 638	Deleted C at 529	Deleted C at 529
qutA4	**Sense primer**	500 to 670 & 556 to 700	Deleted A at 611 Deleted A at 611	3' to 565
qutA384	**Sense primer**	450 to 670 & 463 to 694	none none	3' to 565
	Antisense primer	785 to 1005	none	5' to 694
	I2 central primer	626 to 792 & 650 to 800	Additional C at 766 Additional C at 766	694 to 785
qutA399	**Antisense primer**	778 to 969 & 785 to 1005	none	(843)
	I2 central primer	626 to 792 & 638 to 800	none	
qutA385	**Antisense primer**	778 to 969 & 785 to 1005	Substituted T to G at 812 Substituted T to G at 812	(869)
qutA633	**Antisense primer**	780 to 969 & 800 to 1000	Additional T at 876 Additional T at 876	(869)
qutA379	**Antisense primer**	785 to 1005 & 800 to 1000	Deleted C at 984 Deleted C at 984	(878)

The table shows the extent of sequence obtained from each mutant strain tested in Region 2. Mutations identified and the location of the mutation are givenas identified by DNA sequence and SSCP analysis. The co-ordinates and base alteration refer to the sense strand of the *qutA* ORF. Predicted locationa of mutations are taken from Table 3.4 (in parenthesis) or predicted from previous sequence analysis.

Analysis of the sequence data: The Region 2 antisense PCR primer yields sequence data for Region 2. The physical location of the three mutations identified was entirely consistent with genetic map. The location of the marker strain *qutA633* enables a new set of genetic data to be taken into account when estimating the ratio of bp:RF and therefore should improve the predictive quality of the estimated location of *qutA* mutations. Whilst both the PCR primers had yielded sequence data the extent of sequencing was approximately 250 bp from each primer. The Region 2 sense primer had yielded sequence data from approximately bp 450 to 700, the Region 2 antisense primer yielded sequence data from approximately bp 1000 to 780. Thus the region bp 700 to 780 remains that can not be sequenced from either primer, and it is suspected that two mutations *qutA384* and *399* map to this central region.

Sequence analysis of the centre of Region 2

Design and use of a Region 2 internal sequencing primer: It was decided to design a primer within Region 2 to allow sequencing of this central region. The primer was required to give sequence data between approximately bp 650 and 850 therefore a primer was required in the sense orientation at approximately bp 600, or in the antisense orientation at approximately bp 900. The Region 1B antisense primer mapped at bp 578-599 in the sense orientation. It was decided to use the reverse complementary sequence of the Region 1B antisense primer, since this primer sequence had been previously tested in homology searches. Additionally the primer may be of use as a PCR primer in future work. The Region 2 internal primer (I2) maps to the co-ordinates +578 to +599 (5' CGCATAGAATCTGGAAGCTGCC 3'). I2 was used in sequencing reactions with the Region 2 amplified dsDNA of strains *qutA384* and *399*. Sequencing this central region identified the *qutA384* (Figure 6.1 C) mutation but not the *qutA399* mutation.

Analysis of the sequence data: The location of the *qutA384* mutation identified as the addition of C at bp 767 was consistent with the genetic map. The *qutA399* mutation was not identified and whilst it is possible that the mutation maps 5' of bp 626, it is more likely that the mutation simply was not observed.

Discussion

The order of the five mutations located in Region 2 is consistent with the order predicted from the genetic map and previous sequence data. Again one mutation *qutA399* was not identified and this seems likely to be due to a failure to observe the mutation in the sequencing gels rather than its absence from the region. The *qutA633* mutation mapped to bp 876 within Region 3 and yet this mutation was not identified in SSCP analysis of Region 3.

The location of another of the marker strains (*qutA633*) allowed a new set of genetic data to be used in predicting the location of *qutA* mutations. The argument was reinforced that whilst the order of the genetic map can be accurately determined, low RF values can be misleading when predicting the physical distance between mutations. The predictive quality of the genetic and physical alignment provides only as a very rough guide (within 200 bp) to the location of unidentified mutations (Chapter 7).

Mutant QUTA proteins predicted by *qutA* mutations in Region 2

Five new mutations were identified in Region 2, three are recessive (*qutA385, 633* and *379*) and two semi-dominant (*qutA4* and *384*). The protein sequences predicted by the mutations were determined as described in Chapter 5 and are shown in Table 6.2. This data will be discussed in Chapter 8 when all the available data is compiled.

The sequence data for Region 2 reinforces the major result of the analysis, that no mutations map to the 3' end of the *qutA* gene. This observation maintains interest in the functional significance of putative motifs in the *QUTA* activator protein.

Table 6.2 **Translation of mutant *qutA* sequences identified in Region 2**

Strain and phenotype		Mutation	Alteration to protein	Protein length
qutA4	SD	Deleted A at 611	Frameshift incorporating 2 non-QUTA aa before termination	2 aa added to 205 wt = 207
qutA384	SD	Additional C at 766	Frameshift incorporating 11 non-QUTA aa before termination	11 aa added to 255 wt = 266
qutA385	R	Substituted T to G at 812	Nonsense mutation	270 wt
qutA633	R	Additional T at 876	Frameshift incorporating 7 non-QUTA aa before termination	7 aa added to 292 wt = 299
qutA379	R	Deleted C at 984	Frameshift incorporating 45 non-QUTA aa before termination	45 aa added to 327 wt = 372

The position of the mutations refer to the *qutA* ORF. The amino acid (aa) alterations refer to the amino acid sequence of the wild-type (wt) QUTA protein (Appendix 1). The recessive phenotype is represented by R and the semidominant phenotype by SD.

Chapter Seven

DNA sequence analysis of Region 3 of the *qutA* gene

This chapter describes the sequencing of mutations in Region 3 of the *qutA* gene based upon PCR amplified dsDNA fragments. Region 3 previously defined in Section 3.1, is the 851bp fragment extending from bp726 to 1577 amplified by Primer set 3. The length of the region will require an internal primer to allow sequence analysis of the whole region. SSCP analysis of Region 3 had located only one mutation, *qutA506,* between bp 1102 and 1160. Sequence analysis of Region 2 had identified four mutations (*qutA384, 385, 633* and *379*) that lie within the overlapping sequence between Regions 2 and 3. Due to time limitations, not all mutations mapping to Region 3 were analysed, and only those mutations expected to yield the most informative data were chosen in the final rounds of sequencing.

Preparation of Region 3 sequencing templates: Two of the four mutations previously identified in Region 2 (*qutA633* and *379*) were employed as additional controls for the sequencing method in the analysis of Region 3. Genomic DNA was prepared from the *qutA* mutant strains that map 3' of *qutA385* (Table 3.4) since these are predicted to lie within Region 3. Genomic DNA from each strain (*qutA399, 633, 379 324, 361, 506, 389, 390, 303, 636, 444, 382, 214*) was used as template in PCR amplification with the Region 3 primers, using the reaction conditions previously established (Section 3.4). Sequencing template was prepared from each of the strains and stored at 4^{o}C until sequencing reactions were carried out using the relevant primer.

7.1 Mutations in Regions 3

The first stage in sequencing Region 3 was to confirm mutations *qutA633* and *379* that were previously identified in the 3' end of Region 2. The *qutA399* mutation has not been identified but is predicted to lie to the 5' end of Region 3. Analysis of these strains served three purposes:- Provision of additional controls for the sequencing method. To identify the *qutA399* mutation. To test whether the Region 3 sense PCR primer could be used as a sequencing primer, and if so the extent of sequencing possible.

The length of Region 3 (851 bp) will require an internal primer for complete sequence analysis of the region. Such an internal primer would be placed to yield sequence data from a point

overlapping that obtained from the Region 3 sense primer. Sequencing Region 3 using the sense primer yielded the data summarised in Table 7.1.

The *qutA399* mutation was not identified and may map 5' of bp 785. The mutations in the strains *qutA633* and *379* were each found by sequencing the complementary (sense) strand to that sequenced in the work using the Region 2 primers. In both cases the mutational changes were identical to those previously found (*qutA633* an additional T at 876; *qutA379* a deleted C at 984) thus providing further evidence of the reliability of the sequencing method. The Region 3 sense primer gave sequence data from bp 785 up to approximately bp 1000 in the *qutA* ORF.

Design and use of the internal Region 3 sequencing primer

This internal sequencing primer was required to give sequence data from bp 1000 onwards and was therefore designed around bp 950. The primer sequence was designed by inspection of the published sequence (Beri *et al.,* 1987) and the selected sequence used in the homology searches described at the end of Chapter 4 part II. This search would hopefully remove any possibility of non-specific binding of the primer, which was not expected to be a problem in sequencing, but may allow the primer also to be used as a PCR primer in future work. The search yielded no *Aspergillus nidulans* sequences other than *qutA* with significant identity to the primer. The primer designated Internal Region 3 (I3) is in the sense orientation and maps to the co-ordinates 944 to 962 having the sequence (5' CTTGTGGGCTGTTCCTAGC 3').

Sequencing reaction were carried out using the I3 primer and PCR amplified template dsDNA from the strains *qutA324*, *361*, *506*, *390*, *389*, *303*, *636* and *444* . The first group of sequencing reactions *qutA324*, *361*, *506*, *390* and *389* were carried out using short extension times (1.5min) to obtain data as close to the I3 primer as possible. The second group of reaction *qutA506*, *303*, *636* and *444* were carried out using a longer extension time (7min) to sequence as far from the I3 primer as possible. The I3 primer yielded sequence data from bp 980 to 1335, and the results are summarised in Table 7.1.

Alteration to the published sequence data: Three mutations *qutA303*, *636* and *444* all appeared to contain the same change in sequence, namely GC to CG at bp 1329-30 (the *qutA636* strain contained a further sequence change). The probability of three independently isolated mutants containing the same mutation is very low. Additionally, these three strains cannot be separate isolates of the same mutant. The marker strain *qutA636* was isolated on a different occasion and in a different genetic background to *qutA303* and *444*. The wild-type *qutA*$^+$ strain *R153* was used to provide dsDNA template by PCR amplification of the same region. The sequence data obtained also yielded the same apparent change of GC to CG at 1329-30. Therefore this sequence 'change' is not a mutation but identifies an error in the published sequence. Dr. A. R. Hawkins re-examined the

Table 7.1 **Sequence analysis of Region 3**

Strain	Extent of sequencing		Mutation identified	Predicted /identifed location
qutA399	Sense primer	785 to 1000	none	5' to 785
qutA633	Sense primer	785 to 1000	Additional T at 876	Additional T at 876
qutA379	Sense primer	785 to 1000	Deleted C at 984	Deleted C at 984
qutA324	I3 central primer	980 to 1250	none	1119
qutA361	I3 central primer	980 to 1250	Additional A at 1045	1123
qutA506	I3 central primer	980 to 1250	Additional G at 1119	1131
	I3 central primer	1090 to 1200	Additional G at 1119	
qutA390	I3 central primer	980 to 1300	none (1119 to 1165)	1136
qutA303	I3 central primer	1093 to 1335	none (3' to 1119)	1196
qutA636	I3 central primer	1093 to 1335	Additional T at 1165	1200
qutA444	I3 central primer	1093 to 1335	none (3' to 1165)	1213
qutA382	Antisense primer	1294 to 1541	Subsituted T to A at 1313	1273
	&	1280 to 1450	Subsituted T to A at 1313	
qutA214	Antisense primer	1294 to 1541	Subsituted G to A at 1369	1364
	&	1280 to 1450	Subsituted G to A at 1369	

The table shows the extent of sequence obtained from each mutant strain tested in Region 2., the mutation identified and the location of the mutation as identified by DNA sequence and SSCP analysis. The co-ordinates and base alteration refer to the sense strand of the *qutA* ORF. The predicted location of mutations is taken from Table 3.4. Where no mutation was observed the probable location of the mutation is noted in parenthesis.

original sequencing autoradiographs, and confirmed that the published sequence was mistaken at this point and should read 1328 5' CGC 3' 1330. Translation of the corrected sequence reveals a single amino acid such that a Valine residue (aa 443) is actually a Leucine residue. The residues are both aliphatic amino acids differing by a single methyl group in the aliphatic side chain and thus the difference the correction makes to the structural predictions for the *QUTA* protein are minimal.

Mutations identified in Region 3 using the I3 primer: Resolution of this problem identified three mutations (*qutA506 636* (Figures 7.1 A and B respectively) and *361*) in Region 3 using the I3 primer. Differential labelling in the *qutA361* sequencing lanes meant that whilst the autoradiograph could be analysed no suitable quality image could be produced. The order of the mutations was consistent with the genetic map. SSCP analysis had mapped the *qutA506* mutation to bp 1102 to 1160. Sequence analysis of *qutA506* has identified the mutation as an addition of a G nucleotide at bp 1119 (Figure 7.1 A). Two additional piece of information confirmed the identity of the *qutA506* mutation. First, the mutation altered the sequence 1118 to 1122 to TGAATC creating a *Hinf*I restriction site GANTC by which the mutation was identified in SSCP analysis. Second, new *HinfI* restriction site lies within a 138 bp fragment (1062 to 1200) and forms two fragments 50-60bp and 80-90bp. The new *Hinf*I restriction site in *qutA506* yields two bands of 57 and 81 bp. The genetic map predicts that the *qutA506* and *636* mutations are closely linked. In crosses between these two marker strains and the *qutA* mutant strains the majority of results indicated that *qutA506* lies just 5' of *qutA636,* but a minority indicate that the orientation is reversed. The sequencing results confirm the orientation predicted by the majority of the genetic data.

Sequence analysis of Region 3 using the antisense primer

Due to the limitation of time not all the strains predicted to map to the 3' end of Region 3 could be sequenced. The decision was made to sequence only those mutations expected to yield the most informative data about the functional domains of the *QUTA* activator protein. Sequencing template from the two dominant mutant strains *qutA382* and *214* was prepared for analysis with the Region 3 antisense primer. These strains were chosen, first, because they represented a different and rare phenotype and second, as they lie at the extreme 3' end of the genetic map. Sequence analysis with the Region 3 antisense primer confirmed that the whole of Region 3 could be sequenced with the primers now available. The identity of the two dominant mutations is shown in Table 7.1. The data identified the *qutA382* mutation as a T to A substitution at 1313 (Figure 7.1 C) and the *qutA214* mutation as a G to A substitution at 1369 (Figure 7.1 D). The location of the two mutations is consistent with the order predicted by the genetic map and again were found close to their predicted locations. The order of the genetic map has proved consistent throughout the sequencing

Figure 7.1 **Sequencing autoradiographs of mutation identified in Region 3**

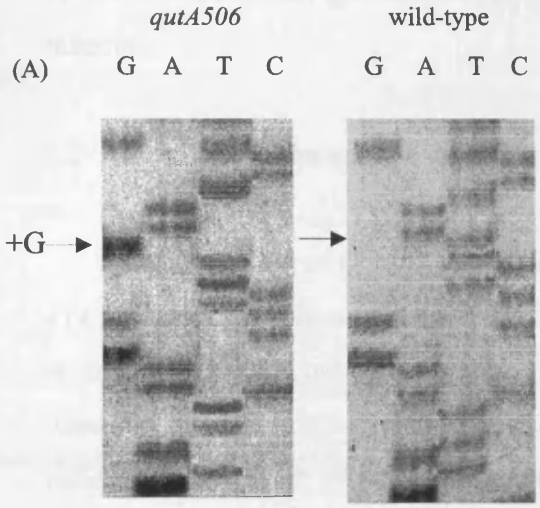


Figure (A) shows the addition of a G in the *qutA506* sequence compared to the wild-type *qutA* gene.

Wild-type 1102 to 1125 ATATTCAAGCGCTCTT AATCCTGT .
qutA506 1102 to 1125 ATATTCAAGCGCTCTTGAATCCTGT

The mutation was identified in Region 3 using the sense primer. The mutation is therefore an additional G at bp 1119.

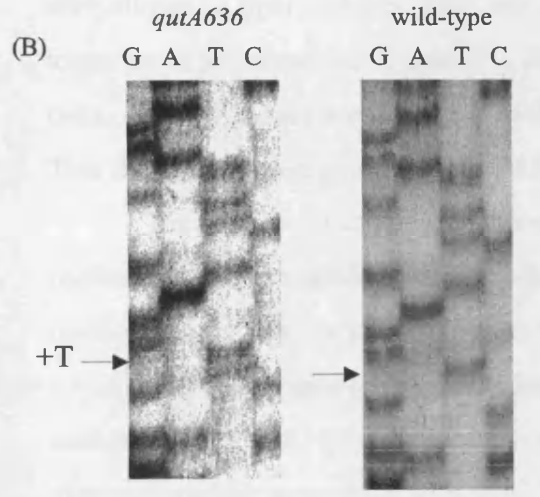


Figure (B) shows the addition of aT in the *qutA636* sequence compared to the wild-type *qutA* gene.

Wild-type 1158 to 1180 GCAGCGT GGATGCTTGTAGGACA .
qutA636 1158 to 1180 GCAGCGTTGGATGCTTGTAGGACA

The mutation was identified in Region 3 using the sense primer. The mutation is therefore an additional T at bp 1165.

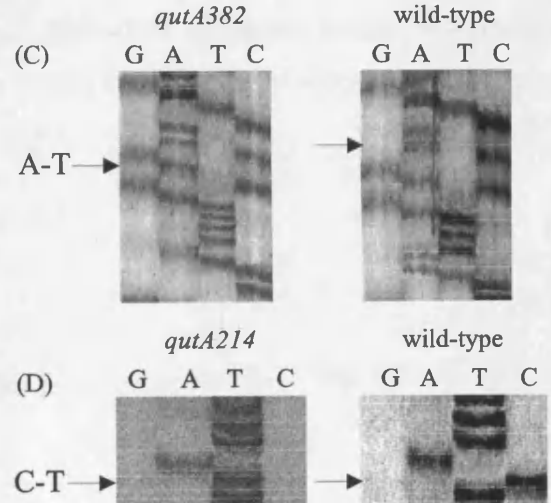


Figure (C) shows the substitution of an A to T in the *qutA382* sequence compared to the wild-type *qutA* gene.

Wild-type 1295 to 1315 CCTAATTTAGCAGCAACTAGA
qutA382 1295 to 1315 CCTATTTTAGCAGCAACTAGA

The mutation was identified in Region 3 using the antisense primer. The mutation is equivalent to a T to A substitution at bp 1313.

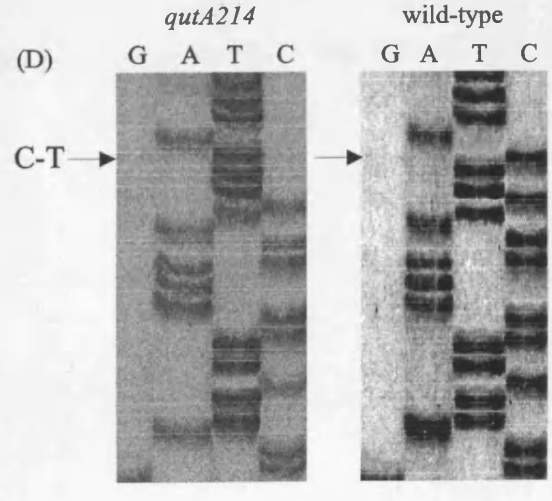


Figure (D) shows the substitution of a C to T in the *qutA214* sequence compared to the wild-type *qutA* gene.

Wild-type 1389 to 1364 GCCATTCTTCCAAACCATCTTCATTT
qutA214 1389 to 1364 GCCATTCTTCCAAACCATCTTTATTT

The mutation was identified in Region 3 using the antisense primer. The mutation is equivalent to a G to A substitution at bp 1369.

of mutants in the *qutA* gene and so I am confidant that the *qutA214* mutation is the extreme 3' mapping.

7.2 Alignment of the genetic and physical maps

The location of three mutations identified and confirmed in Region 3 (*qutA506, 382* and *214*) and a further two mutations identified, but not confirmed (*qutA361* and *636*) were all consistent with the order predicted by the genetic map. It has been noted in previous chapters that the failure to identify a mutation on a sequencing autoradiograph does not necessarily mean that the mutation does not map to the region sequenced. The identification of the remaining marker strains (*qutA506* and *636*) allowed a final revision of the relationship of bp:RF using all the genetic data and actual locations of the mutations (Table 7.2). RF values of less than 10 were not used to calculate the final ratio. Low RF values are unreliable in determining the order of, and distance between, mutations. This final calculation gave the figure 4.3 bp/RF; identical to that obtained originally (Chapter 3.).

The position of each of the mutations identified by DNA sequence analysis and the positions predicted for the remainder of the non-inducible *qutA* mutants is shown in Table 7.3. These data are represented diagrammatically in Figure 7.2.

There is no data which contradicts the hypothesis put forward previously (Chapter 5) that strains *qutA389* and *390* are separate isolates of the same mutation. However there are a number of observations that support the hypothesis: First, the similar genetic map positions and shared phenotype of the two strains. Second, that they are the most 3' mapping semi-dominant mutations and the only pair of semi-dominant mutations falling within approximately 300 bp.

Table 7.2 **Calculation of the bp:RF ratio using all the genetic and physical data**

Strain	*qutA552*		bp:RF	*qutA633*		bp:RF	*qutA506*		bp:RF	*qutA636*		bp:RF
	RF	bp		RF	bp		RF	bp		RF	bp	
qutA402	169	672	**4.0**	292	1120	**3.8**	353	1363	**3.9**	369	1409	**3.8**
qutA422	63	384	**6.1**	186	832	**4.5**	247	1075	**4.4**	263	1121	**4.3**
qutA400	19	80	**4.2**	139	528	**3.8**	200	771	**3.9**	216	817	**3.8**
qutA394	3	64	**x**	123	512	**4.2**	184	755	**4.1**	200	801	**4.0**
qutA552	na			na			na			na		
qutA393	10	68	**6.8**	110	380	**3.5**	171	623	**3.6**	187	669	**3.6**
qutA386	17	101	**5.9**	103	347	**3.4**	164	590	**3.6**	180	636	**3.5**
qutA4	23	185	**8.0**	97	263	**2.7**	158	506	**3.2**	174	552	**3.2**
qutA384	28	339	**12.1**	92	109	**1.2**	153	352	**2.3**	169	398	**2.4**
qutA385	120	384	**3.2**	0	64	**x**	61	307	**5.0**	77	353	**4.6**
qutA633	120	448	**3.7**	na			na			na		
qutA379	122	556	**4.6**	2	108	**x**	59	135	**2.3**	75	181	**2.4**
qutA361	179	617	**3.4**	59	169	**2.9**	2	74	**x**	18	120	**6.7**
qutA506	181	691	**3.8**	61	243	**4.0**	na			na		
qutA636	197	737	**3.7**	77	289	**3.8**	16	46	**3.8**	na		
qutA382	214	885	**4.1**	94	437	**4.6**	33	194	**5.9**	17	148	**8.7**
qutA214	235	941	**4.0**	115	493	**4.3**	54	250	**4.6**	38	204	**5.4**
Individual totals			**77.6/15**			**46.7/13**			**49.7/13**			**56.4/13**
									Overall total 230/54 = 4.3			

The RF values are taken from Appendix 2. The physical distance between mutant sites are the actual distance in bp between mutations identified. na denotes crosses where no RF values are available. x denotes bp:RF ratios where the the RF value is below 10 and consequently the ratio is deemed unreliable.

Discussion

Sequencing of *qutA* Region 3 PCR amplified dsDNA template from *qutA* mutant strains identified five new mutations (*qutA361, 506, 636, 382* and *214*). The position of the mutations were completely consistent with the genetic map. This data allows a final calculation of the relationship of recombination frequency to physical distance using the actual physical locations of all the marker strains and all the genetic data (Table 7.2). The figure together with genetic and sequence data was used to predict the location of the remaining *qutA* mutations not identified by sequencing (Table 7.3).

The genetic map was on the whole an accurate guide to the location of *qutA* mutations, with the only discrepancy coming where there was inconsistency in the genetic data itself. The reliability of the genetic data thus provides confidence in the identification of *qutA214* as the most 3' mapping of the non-inducible *qutA* mutants.

Sequence analysis confirmed that the *qutA506* mutation lay within the region bp 1102 to 1160 resulting in the generation of a new *Hinf*I restriction site by which the mutation had been located by SSCP analysis. Of the ten strains tested by SSCP analysis of Region 3, all bar one (*qutA4*) map within Region 3, and yet not a single alteration in electrophoretic mobility was observed. Of the nine mutations that map to Region 3 five were identified by sequence analysis is (*qutA633, 361, 506, 636* and *214*). This data will be combined with that from Region 1 to allow an evaluation of the SSCP method and factors affecting its effectiveness in identifying mutations (Chapter 8).

Mutant QUTA proteins predicted by *qutA* mutations in Region 3

The mutations in strains *qutA506*, *382* and *214* were identified and confirmed by further sequencing reactions using a different preparation of sequencing template. The mutations in strains *qutA361* and *636* were identified by the analysis of a single preparation of sequencing template. The reliability of the sequencing data attained from identification of other mutations provides confidence in the location of these two mutations. The proteins formed by these mutations were determined as described in Chapter 5 and are shown in Table 7.4. Implications for the function of the *QUTA* activator protein provided by these and previous results are discussed in the next chapter.

Table 7.3 **Predicted and actual ocations of *qutA* mutations**

Strain	Phenotype	Actual location	Predicted location
qutA380	R	(5' of 428)	-756
qutA404	SD	(5' of 428)	-468
qutA402	SD	-244	
qutA422	SD	44	
qutA400	SD	348	
qutA394	SD	364	
qutA552	R	428	
qutA341	SD	(428 to 876)	515
qutA393	R	495	
qutA386	SD	529	
qutA4	SD	611	
qutA384	SD	766	
qutA399	SD	(428 to 876)	829
qutA385	R	812	
qutA633	R	876	
qutA379	R	984	
qutA324	R	(876 to 1369)	
qutA361	R	1045	1018
qutA506	R	1119	
qutA389	SD	(876 to 1369)	1122
qutA390	SD	(5' to 1165)	1123
qutA303	R	(876 to 1369)	1160
qutA636	R	1165	
qutA444	R	(1119 to 1369)	1201
qutA382	D	1313	
qutA214	D	1369	

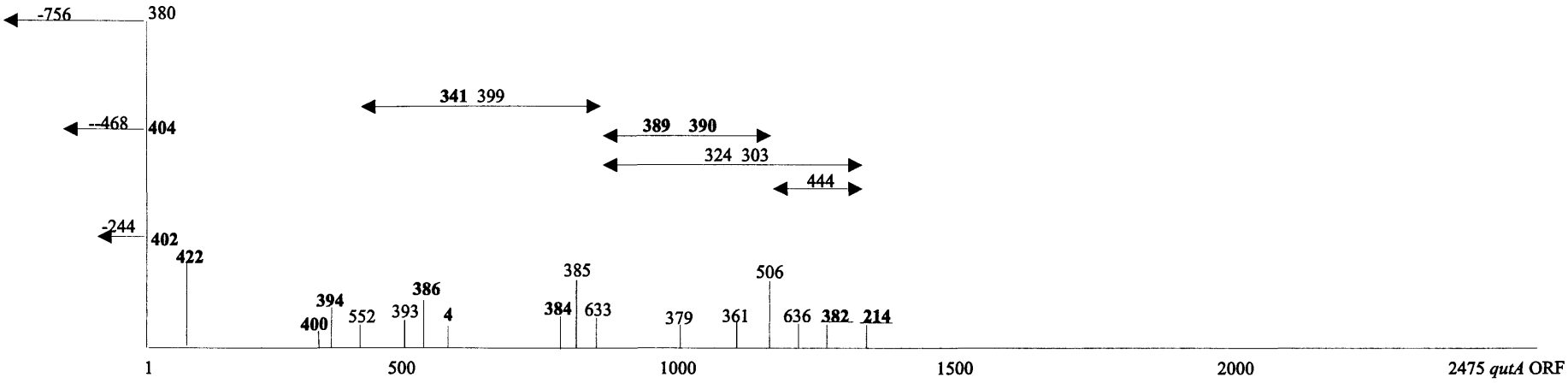
The table shows the location of each of the *qutA* mutations identified. The region where unidentified mutations lie are noted in paraenthesis. Predicted locations for the unidentified mutations are also given based upon RF values and regions sequenced for each mutation.

Table 7.4 **Translation of mutant *qutA* sequences identified in Region 2**

Strain and phenotype		Mutation	Alteration to protein	Protein length
qutA361	R	Additional A at 1045	Frameshift incorporating 48 non-QUTA aa before termination	48 aa added to 348 wt = 396
qutA506	R	Additional G at 1119	Frameshift incorporating 23 non-QUTA aa before termination	23 aa added to 373 wt = 396
qutA636	R	Additional T at 1165	Frameshift incorporating 7 non-QUTA aa before termination	7 aa added to 389 wt = 396
qutA382	D	Substituted T to A at 1313	Nonsense mutation	437 wt
qutA214	D	Substituted G to A at 1369	Single amino acid substitution E to K at aa 457	825

The position of the mutations refer to the *qutA* ORF. The amino acid (aa) alterations refer to the amino acid sequence of the wild-type (wt) QUTA protein (Appendix 1). The recessive phenotype is represented by R and the dominant phenotype by D.

Figure 7.2 **Location of the *qutA* mutations**



The figure shows the predicted location of *qutA* mutations, and is a graphical representation of the data in Table 7.4 The co-ordinates below the line refer to the position in the *qutA* ORF. Strains depicted above marker strains *qutA552, 633, 506 & 636* are genetically to close tro the marker strain to be separated with confidence. Recessive strains are shown in plain text. Semi-dominant strains are shown in bold text and dominant strains in bold underlined text. Strains shown above arrowed lines indicate those mutations not identified by sequence analysis and the regions where they must map.

Chapter Eight

Discussion

The final discussion chapter is presented in eight sections.

The data from single stranded conformational polymorphism (SSCP) analysis will be reviewed with regard to the nature and location of mutations and their detection by the technique.

A review of recent work upon functional motifs in activator proteins, where the presence of similar motifs encoded by the *qutA* gene is discussed and inferences made for the function of the *qutA* activator protein.

Two sections reviewing work by our collaborators, describing the proposed evolutionary origins of the *qutA* activator and a summary of the data obtained by expression of cloned regions of the activator and repressor genes in *E.coli* and *A.nidulans*.

Two sections analysing of the *qutA* mutations identified their distribution and phenotypes attempting to determine the significance of the deduced changes in the activator protein and to explain the phenotype of mutant strains in terms of the molecular action of the mutant protein.

A section combining the data described in this chapter in order to develop a model of the functional motifs and domain structure of the *qutA* activator, and a brief comment on how the developments discussed here relate to work on some other regulatory proteins.

The chapter concludes with a brief summary of the results obtained in the course of this thesis and an outline next stages of work required to further these initial studies on the domain structure and function of the activator protein.

8.1 SSCP analysis as a method of detecting single base mutations

The technique of SSCP analysis was used in the preliminary mapping of *qutA* mutations (Chapter 3). The method identifies mutations by observing the electrophoretic mobility of ssDNA fragments due to differences in their pattern of folding that may result from sequence changes. SSCP analysis was carried out on two regions of the *qutA* gene designated Regions 1 and 3. Three different mutations (*qutA422, 400* and *552*) were detected in Region 1 by their altered electrophoretic mobility (Chapter 3). One mutation was detected in Region 3 by the digestion of a fragment containing a mutation that created a restriction site for the *Hinf*I enzyme that was fortuitously used in the analysis. Therefore in effect the SSCP analysis had identified three mutations *qutA422*, *400* and *552* all in Region 1.

DNA sequence analysis of Region 1 identified one mutation (*qutA394*) tested but not detected by SSCP analysis. Sequence analysis of Region 3 identified five mutations (*qutA633, 361, 506, 636* and *214*) missed by SSCP analysis and showed that four other mutations lying within the region were also missed. To summarise, SSCP analysis had effectively identified three out of a possible four mutations in Region 1 and none out of nine in Region 3. The difference in the success of the technique in the two regions suggests that analysis of the data, to identify which characteristics of a mutation or of the test region may be of value in contributing to detection of mutations by this technique.

The nature of the mutation

The three mutations detected by SSCP analysis involved base addition (*qutA422*), deletion (*qutA400*) and substitution (*qutA552*). Four of the mutations not detected were additions and two were substitutions (Table 8.1). Whilst it is possible that a deletion is more likely to detected than the other two classes this is based on negative evidence, in that none of the undetected mutations are deletions. The actual base involved in the mutation did not seem to have any influence on the chances of a mutation being identified. The three mutations detected all involved an Adenine residue , as did five of the six mutations that were not detected. Neither the type of mutation nor base pair involved, seem to affect the likelihood of detecting a mutation.

The size of the fragment and position of the mutation

The location of the mutation was the next subject for analysis. Two aspects of the location of the mutation were analysed; the length of the fragment within which the mutation was identified and the position of the mutation within that fragment. The length of the fragments in which mutations were detected were compared with the length of the fragments in which the mutation occurred but

Table 8.1 **Summary of SSCP data: Restriction fragments and mutations**

Strain	Restriction fragment and location of mutation			
qutA422	***Hha*I 224 (59*165)**		***Hpa*II 313 (17*296)**	
+A at 44	***Hae*III 150 (96* 54)**		*Alu*I 205 (96*109)	
qutA400	***Hpa*II 145 (8*137)**		***Hae*III 247 (90*109)**	
-A at 348			*Alu*I 320 (163*157)	
qutA552	***Hha*I 250 (173* 77)**		*Hpa*II 145 (83* 57)	
G to A at 428	***Hae*III 247 (170* 77)**		***Alu*I 320 (243* 77)**	
qutA393	*Hha*I 250 (109*141)		*Hpa*II 145 (24*121)	
T to C at 364	*Hae*III 247 (106*141)		*Alu*I 320 (179*141)	
qutA633	*Alu*I 320 (179*141)	*Ava*II 297 (150*147)	*Hha*I 275 (150*125)	
+T at 876	*Hinf*I 336 (150*186)		*Sty*I 145 (74* 15)	
qutA361	*Alu*I 212 (162* 50)	*Ava*II 62 (22* 40)	*Hha*I 112 (44* 68)	
+A at 1045	*Hinf*I 336 (319* 17)		*Sty*I 502 (154*348)	
qutA506	*Alu*I 42 (24* 18)	*Ava*II 171 (34*137)	*Hha*I 43 (6* 37)	
+G at 1119	***Hinf*I 138 (57* 81)**		*Sty*I 502 (228*274)	
qutA636	*Alu*I 440 (28*412)	*Ava*II 171 (80* 91)	*Hha*I 178 (9*169)	
+T at 1165	*Hinf*I 138 (103* 35)		*Sty*I 502 (?*?)	
qutA214	*Alu*I 440 (232*208)	*Ava*II 321 (113*274)	*Hha*I 243 (35*208)	
G to A at 1369	*Hinf*I 133 (76* 57)		*Sty*I 502 (478* 24)	

The table shows the restriction fragments used in SSCP analysis. The location of the mutation within the restriction fragment is noted in paraenthesis. The figures in **bold** are those that were detected by SSCP analysis. The distance from each end of the fragment to the mutation (*) is given.

was not detected (Table 8.2). The data indicate that mutations are most likely to be detected in fragments between 200 and 300 bp in length. The position of the mutations were separated into those that mapped to the centre of the fragment, those that lay close to the ends of a fragment and those that mapped acentrically (Table 8.3). Mutations were more likely to be detected when located acentrically within a fragment. This conclusion is further supported when considering that the two mutations identified as metacentric map to 36% and 64% of the length of the restriction fragment, just within the arbitrary figure chosen to define locations. Additionally two of the acentrically located mutation that failed to show a mobility shift (*qutA361* and *qutA214*) were in restriction fragments over 500 bp long, larger than the optimal size described above.

In summary, when embarking upon SSCP analysis the restriction endonucleases should be selected to produce fragments of optimal length, which in this study were between 200 and 300 bp. Additionally, the sites for the series of restriction enzymes used should occur at intervals such that, any base lying centrally in one fragment is acentrically located in another and terminally located in a third. The findings described above are largely concordant with those obtained by a systematic analysis documenting the sensitivity of the technique in detecting single base substitutions (Sheffield *et al.*, 1993). The study showed that the nature of a substitution mutation played little role in its detection. However, they suggested that addition or deletion mutations may be more readily detected, this is not supported by my findings. In contrast to the finding of this analysis Sheffield *et al.*, proposed that the location of mutation within a fragment was an insignificant factor in determining detection, however alteration of the flanking sequences did increase the likelihood of detection. The optimal size of the DNA fragment for detection of a mutation was reported to be 150 bp (97%) with the detection rate decreasing below 130 bp (70%) and above 200 bp (70%). These data are consistent with other the findings of other workers reviewed by Michaud *et al.*, 1992).

Table 8.2 **Length of DNA fragments to which mutations map**

Fragment length	Detected	Undetected	Rate of detection
1-100 bp	0	4	0%
101-200 bp	2	10	17%
201-300 bp	4	7	36%
301-400 bp	2	5	29%
401-500 bp	0	6	0%

The table shows the classification of restriction fragments by length. The detection or failure to detect a mutation is noted. The final colomn describes the percentage of mutations detected in each restriction fragment length group.

Table 8.3 **Position of mutations within the DNA fragments**

Location	Detected	Undetected	Rate of detection
Metacentric	2	18	10%
Acentric	4	5	44%
Terminal	2	9	18%

The table shows the classification of the position of mutations within restriction fragments. Terminally located mutation lies in the first or last 16% of the fragment length. Metacentirc mutationns are those lying in the central 1/3 of the restriction fragment. Acentric describes mutations lying between 17%-33% and 67-83% of the fragment length. The rate of detection of mutations indicates the optimal position within a fragment for identifiaction of a mutation.

8.2 Functional motifs of activator proteins and the *qutA* gene product

This section considers the recent work on the functional motifs of activator proteins and similar sequences or potential structures found in the *qutA* gene product. Putative functional motifs are therefore identified by association.

The Zinc bi-nuclear cluster DNA binding motif

The zinc cluster DNA binding motif has been identified in more than a dozen fungal regulatory proteins including the *qutA* activator protein. The domain can be considered as three peptide sub-units, the zinc cluster, linker and DNA dependent dimerisation regions (Figure 8.1). The zinc cluster is a compact structure maintained by the chealation of two zinc ions by six cysteine residues (Pan & Coleman, 1990: Vallee *et al.*, 1991: Krulis *et al.*, 1992). The structure of the protein-DNA interaction has been determined for three yeast regulatory proteins *GAL4* (Marmorstein *et al.*, 1992), *LEU3* (Remboutsika & Kohlhaw, 1994) and *HAP1* (Zhang & Guarente, 1994). From these three reports we can see that the proteins bind DNA as a dimer with the zinc cluster fitting into the DNA major groove, where specific contact is made with the central base of the recognition triplet. The *GAL4p* binding site is palindromic suggesting that in the dimer complex with each half of the dimer recognises the same sequence on different strands. The *LEU3p* and *HAPI* dimers bind a directly repeated triplet suggesting a dimer structure with each half of the dimer recognising the same triplet on the same strand (Reece & Ptashne, 1993).

The dimer is stabilised by the DNA dependent dimerisation domain holding the zinc clusters in the correct attitude for stable binding. The dimerisation motif forms an a helical structure common in many protein-protein interactions. The coiled coil structure requires hydrophobic residues at position one and four of every seven amino acid stretch which represents one complete turn of the helix. Proline residues cause a turn in a protein structure and disrupt the helical formation.

Comparison of the *qutA* encoded protein sequence with other activator proteins identified the zinc cluster at amino acids 49 to 76 (Figure 8.1A) and a sequence capable of forming the DNA dependent dimerisation domain at aa 92 to 108 (Figure 8.1B). This analysis also defines the linker region as the sequence between these two domains, that is aa 77 and 91 (Figure 8.1A).

As described earlier the linker region is instrumental in determining the sequence recognised by the DNA binding proteins. The linker region of activators studied shows a more considerable variation in both composition and length compared with the rest of the highly conserved domain. Work by Reece & Ptashne (1993) has shown that these linker regions can be separated into two groups by their function. One set of linker region including *PUT3*, *LEU3*, and *HAP1* bind

Figure 8.1 **Zinc bi-nucleur clusters sub-regions and target sites**

```
A)|          DNA binding          |          Linker          | DNA dependant dimerisation |
                                                              1234567 1234567 1234567
a ACVECRQQKSKCDAHERAPEPCTKC.AKKNVP.C ILKRDFRRTYKRARNEAIEKR FKELTRT.LTNLTSD
b SCTICRKRKVKCDK-LR-P-HCQQCTKTGVAHLC HYMEQTWAEEAEKALLKDNE  LKKLRER VKSLEKT LSKVHSS
c ACLSCRKRHIKCPG-GN-P--CQKC-VTSNAI.C EYLEPSKKIVVSTKY------ LQQLQKD LND
d ACKRCRLKKIKCDQ-EF-P-SCKRC-AKLEVP.C VSLDPATGKDVPRSY------ VFFLEDR LAVMNNN IMA
e ACDICRLKKLKCSK-EK-P-KCAKC-LKNNWE.C RYSPKTKRSPLTRAH------ LTEVESR LERLEQL FLLIF
f ACDACRKKKWKCSK-TV-P-TCTNC-LKYNLD.C VYSPQVVRTPLTRAH------ LTEMENR VAELEQF LKELF
g ACDQCRAAREKCDG-IQ-P-ACFPC-VSQGRS.C TYQASPKKRGVQTGY------ IRTLELA LAWMFEN VAR
h ACDSCRSKKDKCDG-AQ-P-ICSTC-ASLSRP.C TYRANPKKRGLPTGY------ IRELELL WGLVFNK IQG
```

Figure 8.1 A) shows the zinc bi-nucleur cluster motif of eight lower eukaryotic transcriptional activators listed and referenced below. The cluster has been split into three sub-regions.The bold letters in the DNA binding and linker regions indicate homology between at least three out of the eight sequences. Proline residues in the linker region are underlined. The bold letters in the DNA dependant dimerisation sub-region indicate the position of hydrophobic residues necessary to form the α-helical structure.

B) DNA biding sequence | Linker length

LEU3 5'	GCC	(Fixed 6bp)	GGC 3'	21 aa (0 Proline)
HAP1 5'	CGG	(Fixed 6bp)	CGG 3'	20 aa (0 Proline)
PUT3 5'	CGG	(Fixed 6bp)	CCG 3'	12 aa (1 Proline)
PPR1 5'	GCC	(Variable 6bp)	GGC 3'	15 aa (2 Proline)
GAL4 5'	GCC	(Variable 12bp)	GGC 3'	15 aa (2 Proline)
LAC9 5'	GCC	(Variable 12bp)	GGC 3'	15 aa (2 Proline)
QA-1F 5'	GGA	(Fixed 10bp)	TCC 3'	15 aa (1 Proline)
QUTA 5'	GCC	(Unknown assumed fixed 10bp)	NCC 3'	15 aa (2 Proline)

Figure 8.1 B) shows the target sequence for the same transcriptional activators and the relationship between the target sequence and number of proline residues in the linker sequence. The references reporting these sites are also referenced below.

Protein	**Organism**	**Reference**	**Target site evidence & Reference**
a = LEU3	*S.cerevisiae*	Zhou *et al.*, 1987	binding studies Remboutsika & Kohlhaw, 1994
b = HAP1	*S.cerevisiae*	Pfiefer *et al.*, 1989	binding studies Zhang & Guarente, 1994
c = PUT3	*S.cerevisiae*	Marczak & Brandriss 1989	binding studies Reece & Ptashne, 1993
d = PPR1	*S.cerevisiae*	Kammerer *et al.*, 1984	binding studies Reece & Ptashne, 1993
e = GAL4	*S.cerevisiae*	Laughon & Gesteland, 1984	binding studies Reece & Ptashne, 1993
f = LAC9	*S.cerevisiae*	Salmeron & Johnston, 1986	binding studies Halvorsen *et al.*, 1991
g = QA-1F	*N.crassa*	Geever *et al.*, 1989	binding studies Baum *et al.*, 1987
h = QUTA	*A.nidulans*	Beri *et al.*, 1987	promoter motif Hawkins *et al.*, 1988

recognition triplets a constant distance apart; other activators such as *GAL4* and *PPR1* can recognise and bind these triplets despite a difference in the number of bases between them. To be capable of binding triplets separated by a differing numbers bases requires a degree of flexibility as triplets separated by differing numbers of intervening bases will be at different angles within the DNA major groove. Recognition of a directly repeated triplet requires one of the two zinc clusters in the dimer to be rotated by 180 degrees.

The *qutA* activator target binding sequence: The proposed binding site for *QUTA* was determined by means of inspection of the *qutB, D, G,* and *E* promoter regions and identified a 16 bp motif found in the non-coding strand in the promoter of the *qut* regulated genes (Figure 8.2). This consensus sequence is asymmetric suggesting that *qutA* activator does not bind as a dimer or that another protein is involved in DNA binding. However, the identity between the zinc cluster motifs of the *qutA* activator and other DNA binding proteins identified a putative DNA dependent dimerisation subunit, and genetic analysis does not suggest additional proteins involved in *qut* pathway specific transcription activation. Taken together, these observations suggest the activator functions as a dimer.

Comparison of the linker regions of a number of activators (Figure 8.1A) shows that the flexible linker region is associated with the presence of proline residues in the linker, since proline residues are known to disrupt secondary structure formation. The *qutA* activator protein possesses two proline residues in its linker inferring flexibility of binding (Figure 8.1C).

These two factors prompted a revision of the recognition target sequence for the *qutA* activator. Inspection of the promoter regions of *qutA, R, B, D, G, E, C,* and *H* identified a 16 bp motif related to the original promoter motif but with a palindromic arrangement supporting the dimerisation theory. The individual motifs, the revised consensus sequence and the original consensus sequence are shown in Figure 8.2.

Transcriptional activation motifs

The transcriptional activation motif described in the original sequencing of *qutA* at aa 809-824 is an acidic class motif most commonly found in fungal regulatory factors. The motifs show little primary sequence identity, and various structures for the motifs have been predicted including ampiphatic helices and acid blobs (Struhl, 1989: Syler, 1988: Giniger & Ptashne, 1987). The putative motif in the *qutA* activator is a sequence capable of forming the helical structure thought to characterise such domains at the time.

The discovery by Van Hoy *et al.,* (1993) and Leuther *et al.,* (1993) that the motif is not helical but forms a β sheet structure, and that an acidic nature is not essential for its function, prompted a computer aided search for β-sheet structures within the *qutA* activator protein. Using the predicted

Figure 8.2 **Putative activator target motifs in the *qut* cluster**

Gene	Distance	Motif	Match with new consensus
qutA	-500	5' GGCTTGCAGTCGGGGC 3'	9/16
qutB	-124	5' GGCTAAACGCTCCCCG 3'	9/16
qutR	-174	5' GGCAGAAAGGCGTGAC 3'	11/16
qutH	-294	5' GGGATCCGCTTCATCC 3'	11/16
qutC	-139	5' GGGTGAGCATTTTGCC 3'	12/16
qutD	-302	5' GGGCGACGGTGATCCC 3'	12/16
qutG	-181	5' GGCGAACTCTTCCGCG 3'	12/16
qutE	-134	5' GGCAGAGCGTTCTGCC 3'	15/16

New consensus	5' GGCAGACnGTTCTGCC 3'
Old consensus	5' GCCAGAnCGTTCTnCC 3'

The table shows putative QUTA target sites identified in the promoter regions of the *qut* cluster genes. The co-ordinates of the motifs refer to the distance from the underlined G to the ATG start codon of the relevant gene. A consensus sequence provided by Hawkins *et al.*, 1988 and a new consensus sequence including promoter regions from the whole gene cluster are given.

structures programs a number of potential β-sheet structures were identified, and since all known the motifs identified to date are acidic, those β-sheet structures in regions of acidity were considered more probable functional motifs. There are four regions of acidity in the *qutA* encoded activator protein, of these three coincided with β-sheet structures predicted by the Chou Fasman criteria and two with Gorman predictions. The acidic regions and predicted β-sheet regions are shown in Table 8.4.

It is possible that as more data becomes available the acidic class activation motif will be classified into sub-groups, dependent upon the specific adapter used to mediate transcription initiation (Reiger *et al.*, 1993).

Nuclear localisation

Transcriptional activators must be able to enter the nucleus to effect gene expression. The nuclear pore complex (NPC) consists of a central 26nm regulated channel and eight 9nm diameter pores, which allow diffusion of small proteins and ions (Feldherr *et al.*, 1984). The NPC may consist of up to one hundred proteins only about twenty of which have been identified (Reichelt *et al.*, 1990). Nuclear localisation in animals, yeast and presumably in *A.nidulans* is dependant upon cystolic factors. These recognise and bind proteins destined for the nucleus and transport them to the NPC where they are transiently associated with the NPC (Powers & Forbes, 1994: Melchior & Gerace, 1995). The proteins involved in nuclear transport are conserved between yeast and mammalian organisms (Wagner & Hall, 1993) and also in plants, although the cystolic transport components appear to be absent (Hicks & Raikhel, 1995). The transporters recognise non-proteolytic signals encoded by proteins that require nuclear transport. This allows proteins to shuttle in and out of the nucleus. Nuclear localisation signals generally fall into three groups: One, signals with a high arginine, lysine and possibly proline content such as the SV40 type signal (Kalderon *et al.*, 1984) {whose transport is affected by the transporter nucleolin (Xue *et al.*, 1993)}; Two bi-partate signals with two basic regions with a variable distance between them such as is found in the *Xenopus* protein nucleoplasmin (Dingwall *et al.*, 1988); The Third group is has the sequence K,I,P,I,P such as is found in the yeast Matα2 protein (Hall *et al.*, 1984). Finally there are a number of signal sequences which do not fit any of these groups (Dingwall & Laskey, 1991: Siomi & Dreyfuss, 1995). It is known that different proteins require different transporter or chaperone proteins to direct them to the nucleus (Yang & De Franco, 1994). These different chaperone proteins may form the basis of regulating nuclear transport. No nuclear localisation signals have been reported in the *qutA* activator protein and little work has been reported on nuclear targeting of *A.nidulans* proteins. A sequence which bears some similarity to the SV40 large T-antigen nuclear

Table 8.4 **Acidic regions in QUTA and β-sheet structures**

Acidic region	Length	Charge	Chou Fasman	Gorman
151 to 192	39	+10	x	x
448 to 476	28	+5	450 to 460	440 to 460
692 to 718	26	+3	690 to 700	680 to 700
782 to 823	41	+5	780 to 790	x

The table shows peptides regions within QUTA that have an acidic charge and whether or not the region is also predicted to form a β-sheet structure. Two separate methods for predicting β-sheet structure were used. The co-ordinates of peptide regions that are both acidic and form β-sheet structures are given as potential transcriptional activation motifs.

localisation signal forms part of the zinc bi-nuclear cluster motif found in QUTA. The sequence responsible for nuclear localisation of GAL4 lie in the amino terminal 74 amino acids, including the zinc cluster motif, when fused to a β-galactosidase reporter gene the hybrid protein becomes concentrated in the nucleus (Silver *et al.*, 1984). However, the DNA binding and nuclear localisation motifs appear to be functionally distinct and mutations that abolish the ability of *GAL4* protein to locate in the nucleus have been identified (Silver *et al.*, 1986) and *GAL4* proteins that lacks this region are still able to activate gene expression to an extent (Brent & Ptashne, 1985). These observations suggest that more than one domain determines nuclear location. Analysis of the *qutA422* mutation showed that the mutant produces the first 15aa of the wild-type QUTA before an addition alters the reading frame incorporating 23 non-QUTA amino acids. The phenotype of the mutant is semi-dominant this means that either the first 15 aa of QUTA or the 23 fortuitous non-QUTA amino acids must have some affect on *qut* regulation. The chance of 23 effectively random amino acids could affect qut regulation is highly is unlikely. Therefore the affect presumably comes from the first fifteen QUTA aa. The DNA binding, activation and repressor interaction domains have all been accounted for in other regions of the protein. This suggests that perhaps nuclear localisation is at least partially determined by this region. The first 15aa of QUTA show no similarity to the three classes of signal motif identified. If QUTA and the truncated version were transported to the nucleus by a specific transport system, then competition between wild-type and truncated protein may reduce levels of functional QUTA entering the nucleus to a point of physiological significance. Small protein can freely diffuse from the nucleus through the non-regulated 9nm pores in the NPC. The truncated QUTA then diffuse through the NPC and then be picked up and transported inside again exacerbating the competitive affect on nuclear localisation.

8.3 The evolutionary origins of the *qut* activator and repressor proteins

The homology between the *qutR* repressor protein and the carboxyl terminal domains of the AROM protein was discussed in the introduction. The homology suggested that the repressor is able to bind one or more of the quinic acid pathway substrates DHQ and or DHS, which form the inducing signal blocking repression of the activator by the repressor. However, this hypothesis raises the question how the repressor and activator recognise and bind each other? A transcriptional activator is apparently being repressed by a protein derived from an enzyme complex. It has been shown that over-expression of the *qutA* gene in *A.nidulans* leads to constitutive expression of the *qut* structural genes (Lamb, H. K. personal communication). This observation may be interpreted as titration of the repressor by the excess activator, indicating direct stoicheometric interaction of the

activator and repressor, a conclusion supported by work on the homologous regulatory proteins *QA1-F* and *QA1-S* in *N.crassa* (Baum *et al.*, 1987).

The *qutR* encoded repressor is homologous to the carboxyl terminus domains of the *AROM* multifunctional enzyme which is active as a dimer. The enzyme complex is held together by multiple intra and inter-molecular bonds, therefore a protein with homology *AROM* may be expected to bind to the repressor. The *qutA* ORF and *qutR* ORF together are approximately the same length as the entire *aromA* ORF, and a comparison of the *qutA* activator protein sequence and the amino terminus of the *AROM* protein was made (Hawkins *et al.*, 1993).

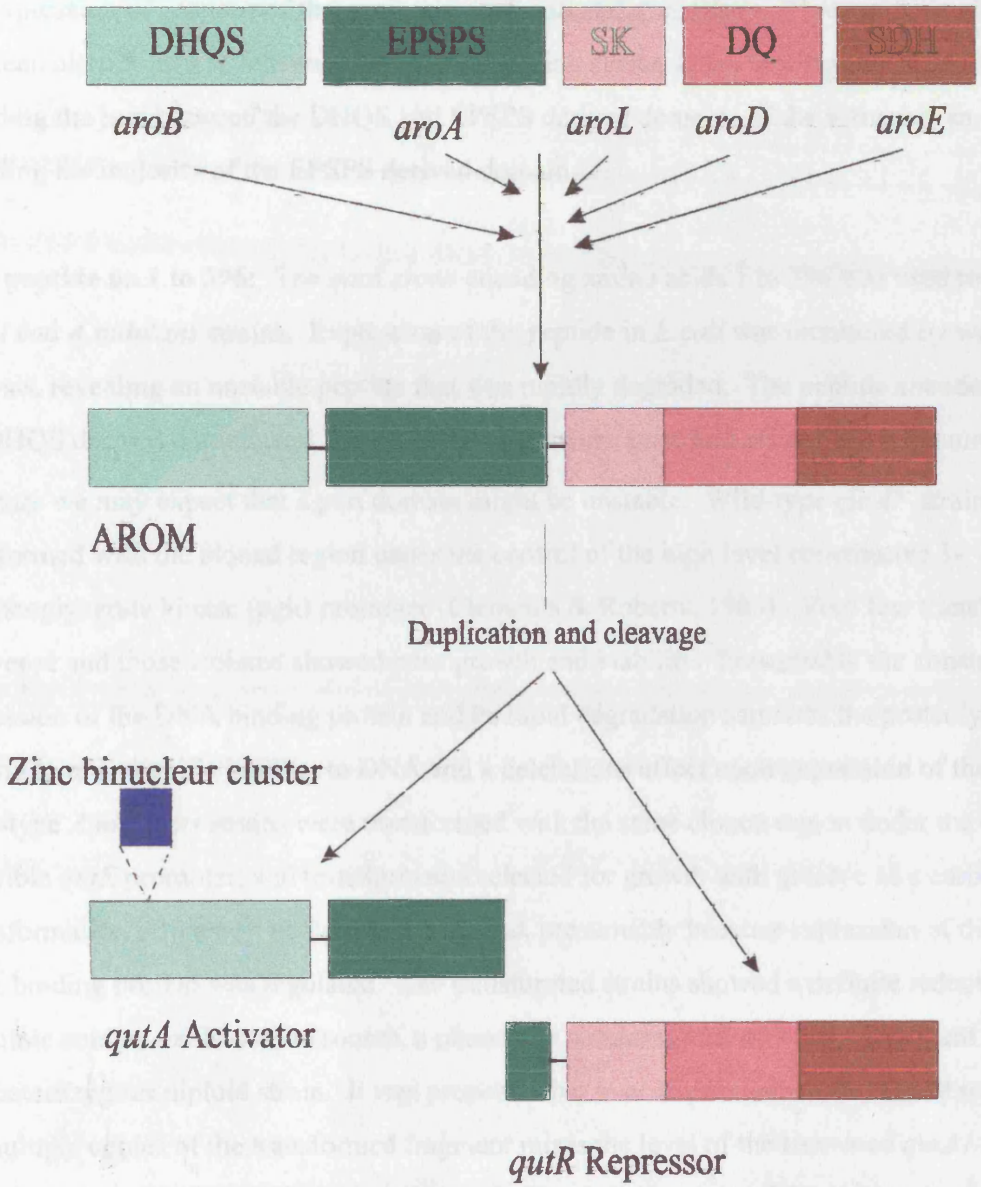
The activator protein shows identity to the two amino terminal domains of the *AROM* protein shown in Figure 8.3 (Hawkins *et al.*, 1993). The homology extends across the whole of the DHQS domain and 80% of the EPSPS domain. The DHQS domain is equivalent to the amino terminus of the *qutA* activator protein (aa 1 to 439) and the EPSPS domain to the carboxyl terminus (aa 440 to 824). The identity is not continuous, and the Zn cluster DNA binding motif appears to have been recruited by insertion. Alternatively, the Zn cluster may have evolved by modification of the DHQS natural Zn binding motif that forms part of the catalytic function of the enzyme. The DHQS and EPSPS domains of *AROM* are able to fold and function independently of the rest of the *AROM* protein (Moore & Hawkins, 1993). Presumably the domains represented in the activator would similarly be able to form a stable protein. More recent work has shown that the DHQS and EPSPS domains can be separated and that the DHQS domain forms functional monomers (Moore *et al.*, 1994).

The whole of the DHQS and 80% of the EPSPS domain from the AROM protein are represented in the activator the remaining 20% of the EPSPS domain and three carboxyl domains are represented in the repressor. This suggests that they were derived from duplication of the *aromA* gene and a single cleavage event (Figure 8.3). This is supported by the observation that both regulatory proteins have diverged from *AROM* by the same extent with the similarity (not identity) being overall 23% with regions as high as 35% at the amino acid level.

Inferences for activator-repressor protein binding: The similarity between the regulatory proteins and the *AROM* pentafunctional polypeptide suggests that the binding sites that hold the *AROM* complex together, particularly in the split EPSPS derived domain, are involved in activator repressor binding. This is discussed further in Section 8.5.

Inferences for *qutA* activator protein structure and function: Both the DHQS and EPSPS domains of the AROM protein can be separated to form stable active monomeric enzymes. The homology between *qutA* and *aromA* predicts a bi-domain structure for the activator protein.

Figure 8.3 **Proposed evolutionary origins of the *aromA* gene and the *qutA* and *qutR* regulatory genes**



The figure shows the proposed evolutionary origins of the pentafunctional *aromA* gene. The figure shows the evolutionary steps that led to the evolution of the *qutA* and *qutR* regulatory genes. The origins of the activator predict a bi-domain structure (Hawkins *et al.*, 1993).

8.4 Expression of cloned regions of the *qutA* gene

During the course of my thesis other work was being carried out in Newcastle on the cloning and expression of portions of the *qutA* gene in *E.coli* and *A.nidulans*. Three regions of the *qutA* gene had been cloned; aa 1 to 396 encoding the entire zinc cluster DNA binding motif; aa 397 to 484 encoding the join between the DHQS and EPSPS derived domains of the activator; aa 485 to 819 encoding the majority of the EPSPS derived domain.

***qutA* peptide aa 1 to 396:** The *qutA* clone encoding amino acids 1 to 396 was used to transform *E.coli* and *A.nidulans* strains. Expression of the peptide in *E.coli* was monitored by western blot analysis, revealing an unstable peptide that was rapidly degraded. The peptide encodes only part of the DHQS derived domain and if the homology between *qutA* and *aromA* has maintained domain structure we may expect that a part domain might be unstable. Wild-type $qutA^+$ strains were transformed with the cloned region under the control of the high level constitutive 3-phosphoglycerate kinase (pgk) promoter (Clements & Roberts, 1985). Very few transformants were recovered and those isolated showed poor growth and viability. Presumably the constant high level expression of the DNA binding protein and its rapid degradation saturates the proteolytic enzymes leading to non-specific binding to DNA and a deleterious effect upon expression of the genome. Wild-type *A.nidulans* strains were transformed with the same cloned region under the control of the inducible *qutE* promoter, and transformants selected for growth with glucose as a carbon source. Transformation efficiency was greatly increased, presumably because expression of the truncated DNA binding protein was regulated. The transformed strains showed a definite reduction in growth on quinic acid as a sole carbon source, a phenotype similar to that of a semi-dominant *qutA* mutation in a heterozygous diploid strain. It was proposed that over-expression of the cloned region caused by multiple copies of the transformed fragment raises the level of the truncated *qutA1-396* protein such that some protein avoids degradation and affects some aspect of the *qut* regulatory circuit (Newton, G. H. personal communication). Possible aspects of *qut* regulation affected in heterozygous diploids with a semi-dominant *qutA* mutation are considered in the next section.
***qutA* peptide aa 397 to 484:** Expression of the peptide aa 397 to 484 in *A.nidulans* had no affect upon *qut* regulation (Lamb, H. K. personal communication).

***qutA* peptide aa 485 to 819:** Transformation of wild-type *qutA A.nidulans* strains with the cloned region encoding the peptide aa 485 to 819 led to the isolation of strains that constitutively express

the *qut* pathway genes. As with over-expression of the entire *qutA* gene this is interpreted as being due to titration of the repressor, but now by a fragment of the activator protein, identifying the region of the activator involved in repressor binding.

***qutR* truncated peptides:** Similar work also underway in Newcastle, cloning and transforming regions of the *qutR* gene into *A.nidulans* have shown that expression of the amino terminal domains of the repressor (excluding the SDH derived domain) are sufficient to reverse the constitutive phenotype of a *qutR16* mutant. These two observations support the hypothesis that the carboxyl terminus of the activator binds the amino terminus of the repressor (Lamb, H. K. personal communication). Further work is currently underway using different cloned regions of *qutR* to determine more with greater accuracy the region of the repressor that binds the activator.

8.5 Distribution of the non-inducible *qutA* mutations: implications for activator structure and function

Sequence analysis of the *qutA* mutant strains had led to the identification of seventeen mutations (Table 8.5 and Figure 8.4). The genetic map proved to be a reliable guide to the order of *qutA* mutations, and the alignment of the genetic and physical maps showed no bias allowing a confident prediction for the location of prior to sequencing *qutA* mutations (Table 3.4). Sequence analysis located sixteen mutations in the *qutA* ORF. The genetic map predicts three mutations in the promoter region of the gene, one of which has been located by sequencing. The sixteen mutations that map in the ORF are located within the first 1369 bp of the 2475 bp gene, distributed on average every 86 bp. The mutations are evenly spread throughout the first half of the gene.

DNA binding domain

It is surprising that no mutations fell in the zinc bi-nuclear cluster DNA binding motif (141 to 324). The original Primer set 1 used in preliminary mapping of *qutA* mutations was designed to cover a region incorporating this sequences as it was thought likely to contain mutations (Chapter 3.1). To explain why no mutations were identified in the zinc bi-nuclear cluster we must consider the way the mutants were isolated.

The *qutA* mutations screened were all selected by filter enrichment following UV mutagenesis (Grant *et al.*, 1988). The technique selects specifically for quinic acid non-utilisation mutants which were then assigned to the respective genes by complementation analysis. The mutant strains are all affected only in their use of quinic acid as a carbon source and are all non-inducible for expression

Table 8.5 **Translation of mutant *qutA* sequences**

Strain and phenotype	Mutation	Alteration to protein	Protein length
qutA402 SD	Additional T at -237	Additional 108 aa before wt QUTA	932
qutA422 SD	Additional A at 44	15 wt frameshift 23 non-QUTA aa	38
qutA400 SD	Deleted A at 348	116 wt frameshift 1 non- QUTA aa	117
qutA394 SD	Substituted T to C at 364	aa substitution I to T at 122	825
qutA552 R	Substituted G to A at 428	Nonsense mutation	142
qutA393 R	Substituted G to T at 495	Nonsense mutation	164
qutA386 SD	Deleted C at 529	177 wt frameshift 30 non- QUTA aa	207
qutA4 SD	Deleted A at 611	205 wt frameshift 2 non- QUTA aa	206
qutA384 SD	Additional C at 766	255 wt frameshift 1 non- QUTA aa	266
qutA385 R	Substituted T to G at 812	Nonsense mutation	270
qutA633 R	Additional T at 876	292 wt frameshift 7 non- QUTA aa	299
qutA379 R	Deleted C at 984	327 wt frameshift 45 non- QUTA aa	372
qutA361 R	Additional A at 1045	348 wt frameshift 48 non- QUTA aa	396
qutA506 R	Additional G at 1119	373 wt frameshift 23 non- QUTA aa	396
qutA636 R	Additional T at 1165	389 wt frameshift 7 non- QUTA aa	396
qutA382 D	Substituted T to A at 1313	Nonsense mutation	437
qutA214 D	Substituted G to A at 1369	Single aa substitution E to K 457	825

The position of the mutations refer to the *qutA* ORF. The amino acid (aa) alterations refer to the amino acid sequence of the wild-type (wt) QUTA protein (Appendix 1). The recessive phenotype is represented by *R* the semidominant by *SD* and the dominant phenotype by *D*.

Figure 8.4 **Putative QUTA proteins produced by *qutA* mutatant alleles**

The figure shows the proteins produced by mutant *qutA* strains as predicted from translation of the mutant sequence The figure shows that all but the promoter mutation and *qutA214* mutation map within the DHQS homologous domain. represents the zinc cluster (aa 49-76). represents the linker region (aa77-95). represents the DNA dependant dimerisation motif (aa95-110). represents the transcriptional activation motif (448-482). repesents the repressor interaction domain (485-819). represents the non-QUTA amino acids (aa) incorporated following a frameshift mutation

of the *qut* cluster genes. The technique of filter enrichment recovers mainly 'non-leaky' mutants that make very little if any growth on quinic acid.

Any mutation that reduced the binding efficiency of the activator protein may simply cause decreased levels of expression of the *qut* cluster genes, yielding partial growth on quinate and thus would have been eliminated by the screening method. Any mutation that removed or altered the specificity of DNA binding may lead to a quinic acid non-utilisation phenotype but may also affect other aspects of the general growth, indeed may be lethal to the mutant strain. Such a mutation may not have been detected by this screening method.

Therefore, whilst it may be surprising that no mutations that completely destroy the function of the DNA binding domain were isolated, a large number of mutations that may occur in this region would not have been isolated by the method described. However this does raise a question about the functional significance of this motif. Work in Dr. A. R. Hawkins laboratory showed that the *qutA* zinc cluster domain does bind zinc (Levesley *et al.*, 1996), and together with the remarkable homology of the motif with other activators strongly supports the argument that this is indeed a functional DNA binding domain.

The 3' half of the *qutA* gene

No mutations were detected in the 3' half of the gene (bp1370 to 2475). The identity between the activator and *AROM* separates the DHQS and EPSPS encoding regions at bp1317 (aa439) of *qutA*. All but one of the *qutA* mutations map in the DHQS derived domain. Mutations must have occurred in the 3' half of the gene. Why were they not recovered? The screening method identifies non-inducible *qutA* mutations, therefore mutations mapping beyond this central region of the gene do not generate non-inducible phenotype. If mutants are inducible they must be capable of activating transcription. Therefore any truncated protein formed by a mutation in the 3' half of the gene must be capable of promoting transcription of the *qut* cluster genes. Two conclusions can be drawn from this deduction.

All the functional domains of the activator protein required for interaction with the target DNA sequences are encoded in the amino terminal half of the protein.

There is a motif in the centre of activator that is capable of acting as a transcriptional activation domain (TAD).

This initial analysis suggests a functionally bi-domain activator protein with the first domain (DHQS derived) domain of the protein capable of DNA interaction and the second domain (EPSPS derived) encoding transcriptional activation and repressor interaction functions.

Restriction mapping sequenced *qutA* mutation

The region that mutations mapped to was used in a restriction mapping exercise using the GCG computer facilities (Devereux *et al.*, 1984). The restriction map obtained was compared to the wild-type *qutA* restriction map. Altered restriction patterns were determined for nine mutations *qutA422, 394, 393, 384, 633, 379, 506, 382, 214* (Table 8.6). New restriction sites were formed for two mutations, the *qutA393* mutation creates a *Dde*I site, the *qutA506* mutation creates a *Hinf*I. The *qutA506* mutation had originally been detected in SSCP analysis by the formation of this additional *Hinf*I site. The *qutA393* mutation also caused the loss of two restriction sites, an *Asu*I site and a *Nla*VI site. The remaining seven mutations caused the loss of a single restriction site.

8.6 Non-inducible *qutA* mutants

There are three classes non-inducible of *qutA* mutant strain dominant, recessive or semi-dominant (Chapter 1). This next section attempts to describe the phenotypes in terms of functions of the activator protein and how the mutant protein might obstruct the normal function of wild-type activator in heterozygous diploids. The final section deals with the promoter mutations and their implications for *qutA* expression and function.

The dominant phenotype

In a diploid containing wild-type activator protein and a dominant mutant *qutA* protein, the *qut* genes are not induced by quinate and there is no growth. The effects of the mutant protein therefore irreversibly prevents the normal action of the wild-type activator protein. If the effects of the mutant protein are permanent and stable the protein itself is presumably stable.

The dominant *qutA382* mutation is a nonsense mutation causing the direct termination of the wild-type activator protein at amino acid 437. The homology between *qutA* and *aromA* marks the boundary of the DHQS domain at amino acid 439. This domain of *AROM* is capable of acting as a separate stable protein. The *qutA382* dominant allele presumably produces a stable protein encoding all the functional domains required for DNA binding, but not transcriptional activation, and the truncated protein binds irreversibly to the promoters of the *qut* regulated genes, strongly supporting the predicted bi-domain structure of the activator protein.

The dominant *qutA214* mutant encodes the most 3' mapping non-inducible mutation identified. The missense mutation cause a single amino acid substitution at aa 456 in the EPSPS derived

Table 8.6 **Restriction map alterations in *qutA* mutations**

Mutation	Restriction site		Region and Restriction pattern
qutA422	*Taq*I	TCGA	Region 1B
(+A bp44)	alteration	TCAGA	wild-type 277, 246, 77, 70, **-48**, 45, 44, 39, **-25**, 22
			qutA422 277, 246, 77, **+73**, 70, 45, 44, 39, 22
qutA394	*Ssp*I	AATATT	Region 1B
(T-C bp364)	alteration	AATACT	wild-type 654, 239
			qutA394 893
qutA393	*Asu*I	GGNCC	Region 1B
(C-T bp495)			wild-type *Asu*I **-372**, 174, **-108**
			*qutA393 Asu*I **+480**, 174
	*Nla*VI	GGNNCC	wild-type *Nla*VI **-372**, 174, **-108**
			*qutA393 Nla*VI **+480**, 174
	*Dde*I	CTNAG	wild-type *Dde*I **-626**, 268
	alteration	GGGACTAAG	*qutA393 Dde*I **+470**, 268, **+155**
qutA384	*Eco*RI*	AATT	Region 2
(+C bp766)	alteration	AACTT	wild-type **-361, -313**
			qutA384 **+674**
qutA633	*Mla*I	TTCGAA	Region 3
(+T bp876)	alteration	TTTCGAA	wild-type **-703, -148**
			qutA633 **+851**
qutA379	*Mnl*I	CCTC(Nx7)	Region 3
(-C bp984)	alteration	C- TC(Nx7)	wild-type **-441,** 166, 66, 62, 61, **-31**, 24
			qutA379 **+472,** 166, 66, 62, 61, 24
qutA506	*Hinf*I	GANTC	Region 3
(+G bp1119)	alteration	GANTC	wild-type 336, **-138**, 134, 133, 93
			qutA506 336, 134, 133, 93, **+81, +57**
qutA382	*Eco*RI*	AATT	Region 3
(T-A bp1313)	alteration	AAAT	wild-type **-544, -230,** 39, 38
			qutA382 **+774,** 39, 38
qutA214	*Mbo*II	GAAGA(Nx8)	Region 3
(A-G bp1369)	alteration	AAAGA(Nx8)	wild-type 377, 179, **-127**, 87, 51, 18 **-12**
			qutA214 377, 179, **+139**, 87, 51, 18

The table shows each mutation that showed an altered restriction map compared to wild-type. The restriction sites are given and the alteration that creates or removes the restiriction site in the mutant strain. The restriction pattern shows the restriction fragments produced by digestion of the relevant region of *qutA*, the fragments are ordered by size in bp. + represent bands in the mutant but not the wild-type. - represents bands in the wild-type but not mutant produced by disruption of a restriction site

domain of the activator. The substitution of glutamic acid for a basic lysine residue destroys transcriptional activation activity. The dominant phenotype is presumably due to the same reasons as described for the *qutA382* allele, that is stable irreversible binding of the transcriptionally inactive activator to the promoter sequences. However whilst the mutation in strain *qutA382* caused the complete removal of the transcriptional activation motif the mutation in strain *qutA214* disables it. Two possible ways the substitution could destroy this activity are described below.

First the mutation occurs near the junction between the two domains, and may cause a conformational change in the structure of the protein such that a spatially distant activation motif is no longer able to interact with the transcription complex.

Second the *qutA214* substitution, that occurs in one of the putative transcriptional activation motifs identified in Table 8.4, may have cause a localised change conformation of the protein at the transcription activation motif. If this latter interpretation is correct then it suggests that the transcriptional activation motif described previously (aa 440 to 460) is the functional motif in the wild-type activator protein.

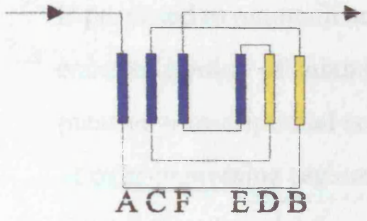
Despite the limited identity between the activator and the *AROM* protein the evidence discussed above suggests that the general bi-domain structure has been maintained, it may be that the general structure within each domain has also been maintained.

The EPSPS derived domain of the activator: The three dimensional structure of the EPSPS protein from *E.coli* (homologous to the EPSPS domain of *AROM*) has been determined by Stallings *et al.,* 1991. The protein forms a complex of six repeated sub-units each approximately seventy amino acids long containing two α-helices and four β-sheets (Figure 8.5A, first part). One face of the a helices and the β-sheets are exposed in each subunit. The sub-units are arranged in two groups of three, joined by a central linking region which includes the ligand binding site (Figure 8.5A, second part). However the six sub-units are not encoded by adjacent peptide sequences. One β-sheet from sub-units one and two are encoded by the amino terminus of the protein, whilst the carboxyl terminus encodes the remainder of these domains in the order subunit two, six then one (Figure 8.5A, third part).

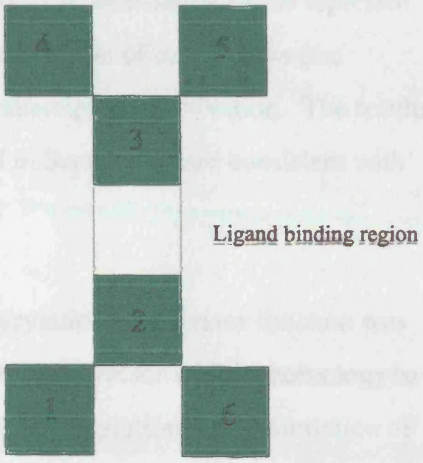
If a similar structure has been maintained in the EPSPS derived domain of the activator and repressor proteins, then knowing the size of each subunit we can formulate a model for activator-repressor binding. The activator encodes the amino terminal 80% of the EPSPS domain. The putative b sheet where the *qutA214* mutation lies corresponds to the b sheet of subunit one encoded by the amino terminus of the *E. coli* EPSPS protein. The carboxyl terminus of the activator is predicted to form a β-sheet and a helix from subunit six, both of these structures are implicated in

Figure 8.5A **The b-sheet, a-helix repeating sub-unit structure of the EPSPS protein**

1) Structure of the EPSPS repeated sub-units



2) Arrangement of the sub-units in the EPSPS proteins



3) Order of the sub-unit components in the EPSPS gene

Amino terminus Carboxy terminus

1	2	3	4	5	3	2	6	1
F	A	B-F	A-F	A-F	A	B-F	A-F	A-D

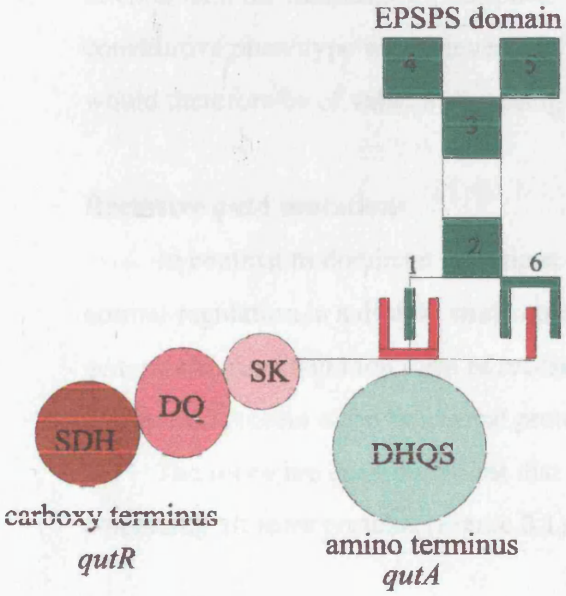
The figure has three parts first the structure of the repeated sub-units that form EPSPS protein each sub-unit encodes approximately 70 amino acids. ▌represents β-sheet regions. ▌represents α-helical regions. The letters A-F, beneath refer to the order that the region occur in an entire subunit.

The second part of the figure shows the arrangement of the sub-units in the entire protein.

The third part of the figure shows the order that sub-unit components are encoded in the protein primary sequence.

The structures refer to the EPSPS protein of *E.coli* (Stallings *et al.*, 1991)

Figure 8.5B **Domain structure of the EPSPS homologous domain separated between the QUTA activator and QUTR repressor proteins**



The figure shows the activator (green) and the repressor (red). Each domain of the activator and repressor is labelled with respect to the homologous domain in the pentafunctional AROM enzyme. The point of interaction between QUTA & QUTR occours at sub-units one and six, each of which is encoded in part by the activator and repressor.
The first β-sheet structure (sub-unit 1) in the EPSPS domain of the activator is the putative transcriptional activation motif.

protein-protein binding. The repressor encodes the carboxyl terminal 20% of the EPSPS peptide including the remaining portion of sub-units six and one. It is the association of these sub-units that is proposed to maintain activator-repressor binding (Figure 8.5B). The association of the repressor encoded portion of subunit one with the activator encoded b sheet portion of subunit one (the putative transcriptional activation motif), is proposed to block transcriptional activation. The results of over-expressing regions of the *qutA* and *qutR* genes discussed in Section 8.3 are consistent with this predicted model for activator-repressor protein binding.

Phosphorylation of the activator protein: The role of phosphorylation in activator function was discussed in the Introduction (Section 1.3). It was proposed that the repressor with its homology to the shikimate kinase domain of *AROM* may be responsible for phosphorylation and potentiation of the activator. This proposal allowed a role for the *qutG* gene product, homologous to bovine myo inositol monophosphatase, as an additional down regulator of the activator. However, both shikimate kinase and inositol monophosphatase recognise metabolites and not proteins raising arguments against this hypothesis. The ligand binding domain of the EPSPS derived domain (the EPSPS synthase domain binds shikimate 3-phosphate, produced by shikimate kinase), is represented within the activator. It may be that like other activators discussed in the Introduction the *qutA* activator protein is potentiated by binding a metabolite itself as well as the effects of the inducer on the repressor, rather than the inducer affecting activator-repressor binding alone. The shikimate kinase derived domain of the repressor may have *in vivo* kinase activity but the substrate may be a metabolite of the *qut* pathway. The metabolite phosphorylated by the repressor may then be bound by the activator at the EPSPS ligand binding motif. EPSPS ligand binding in the *E.coli* enzyme is accompanied by a profound alteration in the conformation of the domain. Ligand binding by the activator may alter the conformation of the domain such that the transcriptional activation motif can interact with the transcription complex. Analysis of sequence changes in the *qutR* gene that lead to a constitutive phenotype would reveal if the mutant protein encodes the shikimate kinase domain and would therefore be of value in assessing the potential in this hypothesis.

Recessive *qutA* mutations

In contrast to dominant mutations the recessive non-inducible mutations have no affect on the normal regulation in a diploid strain containing at least one copy of the wild-type *qutA* gene. In general the most common form of recessive mutation results from total loss of function, that is they produce no protein or no functional protein.

The recessive *qutA* mutations that have been sequenced all occur in the *qutA* ORF and so potentially all form proteins (Figure 8.1). All of these proteins encode the Zn bi-nuclear cluster

DNA binding motif and yet do not interfere with the normal regulation of the *qut* cluster in a diploid. The dominant mutations have been described as stable DNA binding proteins, and this suggests that recessive mutant proteins are either unstable or have a conformation such that they are functionally inactive.

Semi-dominant *qutA* mutations

Heterozygous diploids with a semi-dominant mutant show an intermediate phenotype of clearly visible by reduced growth on quinic acid. A biochemical assay *qut* enzyme levels in these diploids showed ~80% of wild-type diploid enzyme levels (Grant *et al.*, 1988). Dominant and recessive phenotypes can be explained in terms of protein stability and the possession of certain functional domains. The semi-dominant phenotype can be explained in the same terms.

Some semi-dominant mutants potentially produce truncated proteins both longer or shorter than the recessive mutations and some form full length proteins, thus the phenotype cannot be simply explained by the length of the protein or possession of functional motifs. Presumably the truncated proteins encode some scaffolding structure that partially stabilises the mutant peptide, raising the question what functions are encoded by these proteins that affect some aspect of normal *qut* gene regulation.

One interpretation is based on the critical relative concentrations of activator proteins. The most obvious way that the transcriptionally inactive proteins compete with wild-type activator is at the promoter target sites, reducing the overall expression of the *qut* genes and resulting in reduced growth on quinic acid. The intermediate stability of the proteins could allow them to survive degradation long enough to have a deleterious effect on the *qut* expression, similar to the *qutA*(1-396) phenotype described in Section 8.3. However, this interpretation cannot explain the *qutA422* mutation that does not encode the DNA binding motif.

A second interpretation is that the mutant protein may be capable of interacting with the wild-type activator to form an inactive heterodimer, thus diluting the levels of functional activator protein below a threshold required for full induction of the *qut* cluster.

Either of the above interpretations can be applied the *qutA394* mutation that produces a full length activator protein containing a single amino acid substitution of a aliphatic Isoleucine residue for a aliphatic hydroxyl threonine residue at aa 122. The proximity of the mutation to the zinc cluster may affect DNA binding ability dimerisation or destabilise the protein. A possible interpretation applicable to the *qutA422* mutation encoding only the first fifteen wild-type *qutA* amino acids, is that the mutant protein prevents nuclear localisation of the wild-type activator. Entering the nucleus is dependent upon a specific transport system described in Section 8.2. It is possible that the first 15 aa encoded by *qutA* form part of a nuclear localisation motif. The

truncated protein may sequester the components of the transport system, reducing the rate of wild-type activator protein reaching the nucleus below the threshold level required for full expression of the *qut* cluster.

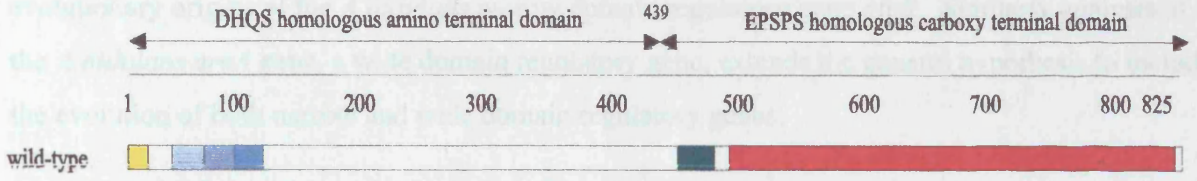
The *qutA422* mutation identifies at least part of a functional motif within the first 15 amino acids of the *qutA* protein. No other function has been ascribed to this region, and the most likely candidate is the nuclear localisation.

Promoter mutations: Three promoter mutations were identified, one of which has been sequenced. A recessive mutation *qutA380*, mapping 5' of bp -265 presumably disrupts some essential promoter feature of the *qutA* gene such that *qutA* is not expressed. Particularly surprising was the fact that two semi-dominant mutations *qutA402* and *qutA404*, maps before the start codon. If the semi-dominant mutations produce no protein or reduced levels of wild-type activator, then the explanations of the semi-dominant and recessive phenotypes would have to be revised. However, it seems more than coincidence that the sequenced *qutA402* mutation should happen to alter a putative ORF such that an ATG start codon mapping at bp -321 to -227 is now in frame with the *qutA* ATG start codon. The mutation therefore allows the potential for an additional 108 aa to be added to the amino terminus of the *qutA* protein. This amino terminal addition might prevent DNA binding, alter conformation or reduce protein stability. For a start codon at bp -324 to be used as an alternate start codon for *qutA* translation, the untranslated leader sequence for the mRNA must be at least 350 bp in length. This would be an unusual feature for *A.nidulans* genes whose average mRNA leader sequence is between 30 and 70 nucleotides (Gurr *et al.*, 1987). If the semi-dominant *qutA404* mutation is identified and shows a similar potential for amino terminal polypeptide addition this would provide additional support for the above prediction. The *qutA* gene has recently been shown to be autoregulated and the size of the mRNA was estimated at to be 3.2 kb compatible with a long leader sequence (Levesley *et al.*, 1996). Taken together, these observations strongly suggest potential for translational control of *qutA* activator expression.

8.7 Functional motifs and domain structure of the *qutA* activator protein

The data presented in this thesis is taken, together with recent results from Dr. A. R. Hawkins' group, to produce a revised model of the functional motifs and domain structure of *qutA* activator protein (Figure 8.6) . The nuclear localisation domain is proposed on the basis of the data for the *qutA422* mutation. The bi-domain structure is based on the homology with *AROM* and is closely correlated with the domain separation indicated by the *qutA382* mutation. The transcriptional

Figure 8.6 **Proposed functional motifs and domain structure of the QUTA activator protein**



The figure shows the location of putative functional motifs and domain structure in the QUTA activator based upon work described in this thesis. Represents the nucleur localisation motif (aa1-15). Represents the zinc binucleur cluster (aa49-76). Represents the specificity or linker region (aa77-90). Represents the dimerisation motif (aa95-110). Represents the transcriptional activation motif (aa448-482). Represents the repressor interaction region (aa485-819). Amino terminus domain DNA binding (aa1-139). Carboxyl terminus domain transcriptional activation and repressor interaction (aa440-825).

activation motif is proposed from the distribution of *qutA* mutations together with the location and nature of the *qutA214* mutation and is concordant with the latest descriptions of the structure of such domains. The region of repressor binding presented is based upon the work by H. K. Lamb described in Section 8.4. Analysis of the three dimensional structure of the EPSPS protein suggests the repressor binding region may be as small as the 50 carboxyl terminal amino acids of the *qutA* protein.

The evolution of regulatory proteins: The proposal for evolution of *qutA* and *qutR* from *aromA* set a precedent for the evolutionary history of other regulatory genes (Hawkins *et al.*, 1992). The evolution of regulatory components from metabolic enzymes recognising similar substrates to the structural components of the pathway they regulate may be a general route for regulator genesis. Work by Hawkins *et al.*, (1994) extended the precedent on the basis of strong evidence for the evolutionary origins of the *A.nidulans* narrow domain regulatory gene *alcR*. Similarly analysis of the *A.nidulans areA* gene, a wide domain regulatory gene, extends the general hypothesis to include the evolution of both narrow and wide domain regulatory genes.

Summary

The proposed homology of the *qutA* activator protein to the two amino terminal domains of the *AROM* protein is supported by the analysis of the *qutA* mutant strains, and **confirms the bi-domain nature of the activator protein**. The amino terminal domain (homologous to DHQS) can bind to the target DNA sequence but has no capacity for transcriptional activation. Comparison of the peptide sequence of the *qutA* zinc bi-nuclear DNA binding motif with homologous motifs proposed a putative dimerisation domain, suggesting the activator binds DNA as a dimer. Analysis of *qut* gene promoter sequences identified a palindromic putative promoter element that **supports the proposal that the activator functions as a dimer**. Perhaps the most significant new finding uncovered by the analysis is that mutations in the *qutA* gene which result in a non-inducible phenotype are located solely in the amino terminal DHQS derived domain and the start of the EPSPS derived domain. It is inferred that mutations lying beyond this domain junction (approximately 50% of the ORF) give rise to a proteins capable of transcriptional activation. A missense mutation in the extreme N-terminus of the EPSPS synthase-like domain (amino acid 457 Glu to Lys) causes a dominant non-inducible phenotype. **Taken together these results identify a potential transcriptional activation motif, however, its position near the centre of the molecule is highly**

unusual. The semi-dominant *qutA422* mutation produces a truncated protein encoding only the first fifteen wild-type amino acids forms, at least part of, a functional motif. No other function has been ascribed to this region and a **nuclear localisation motif is the most likely candidate**. The identification of semi-dominant mutations that map to the *qutA* promoter suggests an unusually long leader sequence with possible **implications for post-transcriptional control of *qutA* expression**.

Further work

The aim of future work is to map out in detail the domain structure of the activator protein and to identify the regions of the protein responsible for transcriptional activation and interaction with the QUTR repressor protein. To achieve this aim there are two main avenues of investigation. Subcloning the entire coding region of the *qutA* gene into an expression vector for production of the entire protein. The subcloning exercise may be repeated using the DNA sequences encoding the two domains of the activator protein separately. The complete activator protein and its constituent domains may be purified in bulk and characterised biochemically by DNA binding studies and cross-linking studies with the DNA target sequence. The native conformation of the complete protein and its constituent domains will be determined by gel filtration chromatography and circular dichroism (CD) measurements. The purified protein would also be used to initiate trials to crystallise the protein. The role of the EPSPS derived domain of the activator may be investigated by construction of site directed mutant forms of the activator protein containing overlapping deletions of the carboxyl terminal domain. These mutant *qutA* genes will be transformed into suitable *qutA*$^-$ mutant host strains of *A.nidulans* shown by DNA sequence analysis to lack functional activator protein. The phenotype of the transformants will be examined for growth on quinic acid and constitutive expression of the qut pathway genes, providing data for location of both the activator-repressor binding site and the transcriptional activation motifs.

Additionally, mapping the transcription start site of *qutA* by S1 nuclease mapping or primer extension is essential to answer the questions posed by the location of the semi-dominant mutations in the *qutA* promoter.

Appendix 1

Appendix 1 shows the nucleotide sequence of the *qutA* gene (Beri *et al.*, 1987) and includes an extended region of promoter sequence Geever (1991 Accesssion number M59936). The adenine residue of the ATG start codon is designated bp +1 and accordingly the promoter region is negatively numbered upto -1 the base prior to the translation start codon. Nucleotide co-ordinates are shown on the left hand side of the page. The protein sequence is shown as a single letter amino acid (aa) abbreviations below the nucleotide sequence. The aa co-ordinates are given to the right hand side of the page. The mutations, in paraenthesis are placed above the appropriate co-ordinates in the sequence. + represents the addition of the subsequent named base after the marked (bold) base in the wild-type sequence. - represents the deletion of the marked (bold) base. * represensts a substitution of the marked (bold) base, the substitution is shwn in the form (wild-type base)-(mutation alteration).

```
-597    ACCCCAG CGAACCAGCA CCGCAAAAAA CCTCCCAATT GAAGTGACGG TTTATGCGGA ATTCAATTGG
-530 ATCGATATGC TTCCCCCGCG GGGAATCTCC GGCTTGCAGT CGGGGCCAAC ATTTGCTGCG CGTCCAGTTC
-460 TACCGTCCAT GTCCACGTTA CCTTAGCTGG GTGCCGGTGG GCAAAAGCGT TCTAGCTCTC TGGCTAGCCT
-390 GACTTAGGAT TTTTACGACA CAAGGTAAGT AACCCATGCT CGAGCGGACG CCTAAAGGTT TCTCATTGTC
-320 ATGGTAGCGG GTAAGCGATA CCGCCGAATT CCCCAACTCC TCTTCTCCGC TCTCTTCCAT CTCCCACAAC
                  +A(qutA402)
-250 ATGATCGCTT TTACCATATG ATCTAATTCA ATTACTTCAT GCATGGTATC TGACTTGGAC GAGTAAGGAT
-180 TTTAACGCTT TGCTCCCGTT GATTAATCGT CGGACATATC CGCGATCCCC AGCTAGGCGG GGCCCTGCAC
-110 ATTCAAACTC CGGGGAAGCA GTTACCAGCC GCCGGAAAAG AATTCATCTC CCTTCCCGTG TCGAAGCTGA
 -40 GCATGTGCTC CGAGCACCTA AGCGCATCAT TCTATCGACT ATGAGTAGCG ATACCCGCCA AACCTCCGGT
                                              M  S  S  D  T  R  Q  T  S  G  10

                +A(qutA422)
  31 GGTAACGCCA GGTCGAAACG GCGATTAACC GATGCGGTCG ACGAGGATGG CAGACCGACC GCCACGGCCG
     G  N  A  R  S  K  R  R  L  T  D  A  V  D  E  D  G  R  P  T  A  T  A  E 34

 101 AAGATCCTAC GTCTAATCCA AAACGCCAAC GAGTCTCTCG AGCCTGCGAT AGCTGCAGAT CGAAGAAGGA
       D  P  T  S  N  P  K  R  Q  R  V  S  R  A  C  D  S  C  R  S  K  K  D  57

 171 TAAATGCGAC GGAGCTCAAC CAATATGCTC GACGTGCGCC TCTCTTTCGC GACCATGTAC CTACAGAGCC
      K  C  D  G  A  Q  P  I  C  S  T  C  A  S  L  S  R  P  C  T  Y  R  A  80

 241 AATCCGAAGA AGCGCGGCCT CCCGACTGGA TACATTCGCA CGTTGGAGTT GCTATGGGGG CTTGTGTTCA
     N  P  K  K  R  G  L  P  T  G  Y  I  R  T  L  E  L  L  W  G  L  V  F  N 104

                                          -(qutA400)        *T-C(qutA394)
 311 ATAAGATACA GGGCAGTGAG GAGGTCGTCC GGACATTATT GAGGGCGGCA AATATTCCCA GCCACCTGGC
       K  I  Q  G  S  E  E  V  V  R  T  L  L  R  A  A  N  I  P  S  H  L  A  127

                                                  *G-A(qutA552)
 381 AACCATGGGG AAAGAGTCGG AGGGGTCTGA TACACTGCTC TCGTATTGGA AAAACAGCAT TGTCTTGAAA
      T  M  G  K  E  S  E  G  S  D  T  L  L  S  S  W  K  N  S  I  V  L  K  150

                                               *C-A(qutA393)
 451 GAGATTGAAC GTCTCTTGAC GTTTCTCGAG CAACCGGAAG GGGACCAAGA AAGAAGCGCG AGAGGAGAAA
     E  I  E  R  L  L  T  F  L  E  Q  P  E  G  D  Q  E  R  S  A  R  G  E  I 174

            -(qutA386)
 521 TCGATTCCCC AGCGGATGCT GAAGAAAGCA GTGTGCTTTC TCCGGAGACG CTTGAGTGGC AGCTTCCAGA
       D  S  P  A  D  A  E  E  S  S  V  L  S  P  E  T  L  E  W  Q  L  P  D  197

                         -(qutA4)
 591 TTCTATCGCG GTTGCGAGTC AAAGCCCGTT AGCGTCTGGG CCATCCCCAG TGAGGCTGCC TCGGCCATCC
      S  I  A  V  A  S  Q  S  P  L  A  S  G  P  S  P  V  R  L  P  R  P  S  220
```

661 ACAACCCGTC TCGTCAGGGA CTCTGGCACG CAAACGATTC CCCTAGGAGA AATCGAAGAC CTTACCACTA
T T R L V R D S G T Q T I P L G E I E D L T T N 244

+C (*qutA384*)

731 ACGGATCATC CCACGATCGG CCTGTGATCG CTTCC**A**ATTC TGCCCGAGAA GAACACCGCT TGCCTCCTAA
G S S H D R P V I A S N S A R E E H R L P P N 267

*T-G (*qutA385*)

801 CCCTTGGCCT T**T**ACTTGATA TATACTTTTC CTACACGCAA TGTTGGTTCC CAATATTAGA AAAACACGAT
P W P L L D I Y F S Y T Q C W F P I L E K H D 290

+T (*qutA633*)

871 ATTCT**T**CGAA CAGCTTTCCG CCAAGGTGAC GATGACCAAT ACAACTCTCC ATCCGCCGCA GGGGACAACG
I L R T A F R Q G D D D Q Y N S P S A A G D N A 314

-(*qutA379*)

941 CGGCCTTGTG GGCTGTTCTA GCACTTGCCT CAATCCAACA AAC**C**TCAATA TCCACTACGC GCCAGTTATC
A L W A V L A L A S I Q Q T S I S T T R Q L S 337

+A (*qutA361*)

1011 AGACTTACCA GAGGACCGAC CTGATCCTGA TCAA**C**TATAT GCCAAAGCAC GGAGTCTTAT CCCAACAGAG
D L P E D R P D P D Q L Y A K A R S L I P T E 360

+G (*qutA506*)

1081 AGTGGGACCT ACCAGCTCGG CCATATTCAA GCGCTCTT**A**A TCCTGTCATT GATAAAGCTT GGCCAGCAGG
S G T Y Q L G H I Q A L L I L S L I K L G Q Q D 384

+T (*qutA636*)

1151 ACTGCGCAGC AGCG**T**GGATG CTTGTAGGAC AGGCTGTTAG GTCTGCCCAG AGTCTAGGGT TAAATGATCC
C A A A W M L V G Q A V R S A Q S L G L N D P 407

1221 TTCCGATGCT ACGGGAGTAG AGAAGACCGC TGGACGGTCC AAACATGTTT TTCTAGGGTG CTTTGTGCTG
S D A T G V E K T A G R S K H V F L G C F V L 430

*T-A (*qutA382*) **GC** (Correction)

1291 GAGACTCTAG TTGCTGCTAA AT**T**AGGCCTG TTGCCTTCCG TGCGCAAGAC TGACTTGACA AAGGTTGGGT
E T L V A A K L G L L P S V R K T D L T K V G L 454

L (Correction)

*G-A (*qutA214*)

1361 TAATAAAT**G**A AGATGGTTTG GAAGAATGGC ACCCTTGGGA AGACCAGACT GGACTTCGCC CGATAGAATC
I N E D G L E E W H P W E D Q T G L R P I E S 477

1431 ATCCCGCTCT TTCCAACGCG GGCCCCTCCA TGCTCTCAGC ACTTTCAACC GTCTGCTCTC GCTAATGTGC
S R S F Q R G P L H A L S T F N R L L S L M C 500

1501 ATTCTGAATG AGTTATGTTG CGTGAGGCAA ACACCAGCGA ATTCCATGTC ATATCTAGGG ACTCTTGAGC
I L N E L C C V R Q T P A N S M S Y L G T L E R 524

1571 GACAACTGCA GCTTTGGGTC TCGGCACTTC CTACAAGTTA TCGGATAGAC CTGCAGACAG TCTCGACCAA
Q L Q L W V S A L P T S Y R I D L Q T V S T K 547

1641 GCCGGCATCA CCCCACATAT TCGGCCTAGA GATGATGTAC GAGGCCGTTG CAACCGCTAT AAGCTTACAG
P A S P H I F G L E M M Y E A V A T A I S L Q 570

1711 ATGGCAGTCC AGAAAAACGA AAGAAACGGA TTACGACGTG CTTCGGAAGG CTCCAAGCGG TTGGTCTCAC
L A V Q K N E R N G L R R A S E G S K R L V S L 594

1781 TACTTCAAGC CTACATGGAG ACCTATAGCC TTTCTGCTAC CTGCCCGACT TTCGGTATAG CTCTGACTTT
L Q A Y M E T Y S L S A T C P T F G I A L T L 617

1851 AGGCCTTCCA CTTCCCTGCA TGAAGAAAAC TGCCCCATGG CCCTTTGAGG CTTCTCATGG CATCAACCAC
G L P L P C M K K T A P W P F E A S H G I N H 640

1921 AAACTCCAAT CGTTTTCTGC CCATCTCGCA ACGGTTTGGA TCCATAATAC TGGCTCGCAG AGAAGACCGA
K L Q S F S A H L A T V W I H N T G S Q R R P R 664

1991 GACACGCTAC ACAGCAAGGT ACCCATCCAA GGAGTCCGAT GGCCATAAGC CTGCCCGGAA ACAATATGCG

```
      H  A  T  Q  Q  G  T  H  P  R  S  P  M  A  I  S  L  P  G  N  N  M  R  687

2061 TTTGACGAAC TTAATCGAAG AATCTCGAAC TGGGAATCTT TCAACCACTG ACTCATATCT TTCGCCAACT
       L  T  N  L  I  E  E  S  R  T  G  N  L  S  T  T  D  S  Y  L  S  P  T  710

2131 TGGATGCGGA CTTCGAACGA TGAAAACGCT GTATTATCTC TACCGACGCC AGCATCGTCT CTTAATATCG
     W  M  R  T  S  N  D  E  N  A  V  L  S  L  P  T  P  A  S  S  L  N  I  A 734

2201 CGTCTGGAGT TGAAACGAAC CCTACTTCGC AGCAAATCAC CCATCAACAT CGCTCTTCGG TCGTTGGAAA
       S  G  V  E  T  N  P  T  S  Q  Q  I  T  H  Q  H  R  S  S  V  V  G  K  757

2271 GCCTAACTCA GCACTGATAA TGTCCGACCT GACGACGCCT TTTCCTGCGT CCGGTCATCA CTATCAACAA
       P  N  S  A  L  I  M  S  D  L  T  T  P  F  P  A  S  G  H  H  Y  Q  Q  780

2341 ACGTATGATG ATGCCACTTT GCACTTCAAT TCTTTAGCCG ACATCCAGAG CACTAGCTCA GCACAACGAC
     T  Y  D  D  A  S  L  H  F  N  S  L  A  D  I  Q  S  T  S  S  A  Q  R  P 804

2411 CAAGGATTGC TCCTGACCTC GATGCCTTGT TTGACGAGTT GGCCTCGCTT GATGGGACAG ATAGGTAACC
       R  I  A  P  D  L  D  A  L  F  D  E  L  A  S  L  D  G  T  D  R  *     825

2481 AAGCTCTCTT CTGCACGCAA GTTATAAAAC TAACTAGCAC TAGGATGGAC AACCACCCCG AATTC
```

Appendix Two

Table 1 **Genetic mapping of the *qutA* activator gene locus**

Strain & Phenotype	*qutA552* (R)	*qutA633* (R)	*qutA506* (R)	*qutA636* (R)
***qutA380* (R)**	417<|	x	436<|	x
***qutA404* (SD)**	133<|	x	369<|	x
***qutA402* (SD)**	172<|	x	242<|	x
***qutA423* (SD)**	70<|	x	458<|*	x
***qutA422* (SD)**	63<|	x	232<|	x
***qutA400* (SD)**	19<|	x	174<|	x
***qutA394* (SD)**	(3)	88<|	245<|	106<|*
***qutA341* (SD)**	|> 6	84<|	x	165<|
***qutA393* (R)**	|> 10	136<|	193<|	214<|
***qutA386* (SD)**	|> 17	34<|	151<|	55<|*
***qutA4* (SD)**	|> 23	22<|	99<|	127<|
***qutA384* (SD)**	|> 28	21<|	83<|	117<|
***qutA399* (R)**	|>114	11<|	47<|	64<|
***qutA385* (R)**	|>104	(0)	41<|	10<|
***qutA379* (R)**	|>141	(2)	32<|	18<|
***qutA324* (R)**	|>371	|> 33	(3)	x
***qutA361* (R)**	|>163	|> 6	(2)	5<|
***qutA389* (SD)**	x	x	(1)	10<|
***qutA390* (SD)**	*|>868	|> 20	(2)	7<|
***qutA303* (R)**	|>102	|> 66	x	(1)
***qutA444* (R)**	x	|> 39	|> 19	x
***qutA382* (D)**	x	|>132	|> 23	|> 17
***qutA214* (D)**	|>286	|>221	|>144	|> 38

The table shows the frequency (per million ascospores plated on selective media) of QA utilising recombinants (*qut*$^+$) in crosses between the *qutA* mutants indicted. Arrows indicate which side of the double mutant marker strain the particular alleles are located, thereby defining the interval to which the mutation was allocated based upon the distribution of the *qutR16* flanking marker. Numbers in paraenthesis indicate very close linkage where the mutation could not be oredered with confidence. An asterix (*) marks inconsistencies within the genetic data. Diploids with the wild-type *qutA* allele produce three phenotypes: Dominant *qutA* alleles (D) produce no growth; recessive alleles (R) growth identical to a *qutA*$^+$ homozygote; semidominant *qutA* alleles (SD) show a marked decrease in mycelial density and conidiation (approximately 75% of the wild type homozygotous diploid).

Table 2 **Genetic distance between double mutant *qutA* alleles**

Strain	*552*	***552* to *633***	*633*	***633* to *506***	*506*	***506* to *636***	*636*
qutA380	<417	**x**	x	**x**	<436	**x**	x
qutA404	<133	**x**	x	**x**	<369	**x**	x
qutA402	<172	**x**	x	**x**	<242	**x**	x
qutA423	< 70	**x**	x	**x**	<458	**x**	x
qutA422	< 63	**x**	x	**x**	<232	**x**	x
qutA400	< 19	**x**	x	**x**	<174	**x**	x
qutA394	<3>	**85**	<88	**157**	<245	**-139***	<106
qutA341	6>	**90**	<84	**x**	x	**x**	<165
qutA393	10>	**146**	<136	**57**	<193	**21**	<214
qutA386	17>	**51**	<34	**117**	<151	**-96***	<55
qutA4	23>	**45**	<22	**77**	<99	**28**	<127
qutA384	28>	**49**	<21	**62**	<83	**34**	<117
qutA399	114>	**125**	<11	**36**	<47	**17**	<64
qutA385	104>	**104**	<0>	**41**	<41	**-31***	<10
qutA379	141>	**139**	<2>	**30**	<32	**-14***	<18
qutA324	371>	**338**	33>	**36**	<3>	**x**	x
qutA361	163>	**157**	6>	**8**	<2>	**7**	<5
qutA389	x	**x**	x	**x**	<1>	**8**	<10
qutA390	*868>	**848***	20>	**21**	<2>	**6**	<7
qutA303	102>	**102**	66>	**x**	x	**x**	<1>
qutA444	x	**179**	39>	**20**	19>	**x**	x
qutA382	x	**x**	132>	**109**	23>	**6**	17>
qutA214	286>	**65**	221>	**77**	144>	**106***	38>
	Av 1559/13 = **120**		Av 848/14 = **61**		Av 233/9 = **16**		

The table shows the average distance between the marker alleles used to construct a compilation table of the range of genetic distance across the *qutA* locus.

Table 3 **Genetic distance betwee *qutA* alleles used to complie Figure 1.6**

Strain	*552*	*633*	*506*	*636*	Total
qutA380	417				**0**
qutA404	172				**245**
qutA402	133				**284**
qutA423	70				**347**
qutA422	63				**354**
qutA400	19				**398**
qutA394	<3>				**414/420**
qutA552	x				**417**
qutA341	6				**423**
qutA393	10				**427**
qutA386	17				**434**
qutA4	23				**440**
qutA384	28				**445**
qutA385	104				**521**
qutA399	114				**531**
qutA633	x				**537**
qutA379		<2>			**539**
qutA324			<3>		**595/601**
qutA361			<2>		**596/600**
quta506			x		**598**
qutA390			<1>		**597/599**
qutA389			<2>		**596/600**
qutA303				<1>	**613/615**
qutA636				x	**614**
qutA444			19>		**617**
qutA382				17	**631**
qutA214				38	**652**

The data shown in the table above was selected from the previous tables 1 and 2 to construct a genetic map spanning the entire *qutA* locus.

Literature cited

Adams, B. G. & Dalbec, J. M. (1977) Non pleiotropic nature of the *gal3* mutation in yeast. *Biochem. Biophys. Res. Comm.* **7**, 1348-1354.

Adrianopolous, A. & Hynes, M. J. (1990) Sequence and functional analysis of the positively acting regulatory gene *amdR* from *Aspergillus nidulans*. *Mol. Cell. Biol.* **10**, 3194-3203.

Agre, P., Johnston, P. F. & McKnight, S. L. (1989) Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. *Science* **246**, 922-925.

Anton, I. A., Duncan, K. & Coggins, J. R. (1987) A eukaryotic repressor protein, the *qa-1S* gene product of *Neurospora crassa*, is homologous to part of the AROM multifunctional enzyme. *J. Mol. Biol.* **197**, 367-371.

Aramayo, R. & Timberlake, W. E. (1990) Sequence and molecular structure of the *Aspergillus nidulans yA* (laccase- I) gene. *Nucl. Acids Res.* **78**, 3415.

Aramayo, R. & Timberlake, W. E. (1993) The *Aspergillus nidulans yA* gene is regulated by *abaA*. *EMBO J.* **12**, 2039-2048.

Ardnt, K. &. Fink, G. R. (1986) GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5' TGACTC 3' sequences. *Proc. Natl. Acad. Sci. USA* **83**, 8516-8520.

Armitt, S., McCullough, W. & Roberts C. F. (1976) Analysis of acetate non-utilising (*acu*) mutants in *Aspergillus nidulans*. *J. Gen. Microbiol.* **92**, 263-282.

Arst, H. N. &. Cove, D. J. (1973) Nitrogen metabolite repression in *Aspergillus nidulans*. *Mol. Gen. Genet.* **126**, 111-141.

Arst, H. N. & MacDonald, D. W. (1973) A mutant of *Aspergillus nidulans* lacking NADP-linked glutamate dehydrogenase. *Mol. Gen. Genet.* **122**, 261-265.

Arst, H. N., Brownlee, A. G. & Cousen, S. A. (1982) Nitrogen metabolite repression in *Aspergillus nidulans* - A farewell to *tamA*. *Curr. Genet.* **6**, 245-257.

Arst, H. N. &. Scazzochio, C. (1975) Initiator constitutive mutation with an up-promoter effect in *Aspergillus nidulans*. *Nature* **254**, 31-34.

Arst, H. N. &. Scazzochio, C. (1985) Formal genetics and molecular biology of the control of gene expression in *Aspergillus nidulans*. *In Gene manipulations in Fungi eds Bennet, J. W. & Lasure, L. L. Academic Press, Orlando*. pp309-343.

Arst, H. N., Tollervey, D. & Caddick, M. X. (1989) A translocation associated, loss-of-function mutation in the nitrogen metabolite repression regulatory gene of *Aspergillus nidulans* can revert intracistronically. *Mol. Gen. Genet.* **215**, 364-367.

Arst, N. H. (1990) Wide domain regulation of gene expression in *Aspergillus nidulans* with special attention to the role of the *meaB* gene. In: Haslot, H., Davies, J., Florent, J., Bobichan, L., Durand, G., and Penasse, L. (Eds). Proc. 6th Int. Symp. Genet. Ind. Microorg. GIM90. Paris. Soc. Fr. Microbiol. 2 (1990), pp555-566.

Axelrod, J. D., Major, J. & Brandriss, M. C. (1991) Proline independant binding of *PUT3* transcriptional activator protein detected by 'footprinting' *in vivo*. *Mol. Cell. Biol.* **11**, 564-567.

Bachmann, B., Luke, W. & Hunsmann, G. (1990) Improvement of PCR amplified DNA sequencing with the addition of detergents. *Nucl. Acids Res.* **18**, 1309.

Bai, Y. & Kohlaw, G. B. (1991) Manipulation of the zinc cluster region of transcriptional activator *LEU3* by site-directed mutagenesis. *Nucl. Acids Res.* **19**, 5991-5997.

Bailey , C. &. Arst, H. N. (1975) Carbon catabolite repression in *Aspergillus nidulans*. *Eur. J. Biochem.* **51**, 573-577.

Baum, J. A., Geever, R. & Giles, N. H. (1987) Expression of *qa-1F* activator protein: Identification of upstream binding sites in the *qa* gene cluster and localization of the DNA binding domain. *Mol. Cell. Biol.* 7, 1256-1266.

Bentley, R. (1990) The shikimate pathway -a metabolic tree with many branches. *Critical Reviews in Biochemistry and Molecular Biology* **25**, 307-384.

Beri, R. K., Whittington, H., Roberts, C. F., & Hawkins, A. R. (1987) Isolation and characterization of the positively acting regulatory gene *qutA* from *Aspergillus nidulans*. *Nucl. Acids Res.* **15**, 7991-8001.

Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B. L., Goodfellow, P. N. & Fellous, M. (1990) Genetic evidence equating SRY and the tetis-determining factor. *Nature*. **348**, 448-550.

Birnboim, H. C. & Doly, J. (1979) A rapid alkalkine extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* 7, 1513-1523.

Bram, R., J. & Kornberg, R. D. (1985) Specific protein binding to far upstream activating sequences in polymerase II promoters. *Proc. Natl. Acad. Sci. USA* **82**, 43-47.

Brandriss, M. C. & Magasanik, B. (1979) Genetics and physiology of proline utilisation in *Saccharomyces cerevisiae*: enzyme induction by proline. *J. Bacteriol.* **140**, 498-503.

Brandriss, M. C. (1987) Evidence for positive regulation of the proline utilisation pathway in *Saccharomyces cerevisiae*. *Genetics* **117**, 429-435.

Brent, R. & Ptashne, M. (1985) A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **43**, 729-736.

Brisco, P. R. G. & Kohlaw. G. B. (1990) Regulation of *LEU2*: total deletion of regulatory gene *LEU3* unmasks *GCN4* dependant basal level expression of *LEU2*. *J. Biol. Chem.* **265**, 11667-11675.

Broach, R. J. (1979) Galactose regulation in *Saccharomyces cerevisiae*. The enzymes encoded by the *GAL7, 10,1* clusters are co-ordinately controlled and separately translated. *J. Mol. Biol.* **131**, 41-53.

Caddick, M. X., Arst, N. H., Taylor, L. H., Johnston, R. L. & Brownlee, A. G. (1986 a) Cloning of the regulatory gene *areA* mediating nitrogen metabolite repression in *Aspergillus nidulans*. *EMBO J.* **5**, 1087-1090.

Caddick, M. X., Brownlee, A.G. & Arst, N. H. (1986 b) Phosphate regulation in *Aspergillus nidulans*: responses to nutritional starvation. *Genet. Res. Camb.* **47**, 93-102.

Cariotis, M., Jones, R. F. & Wessling, A. C. (1974 a) Cross pathway regulation: Histidine-mediated control of histidine, tryptophan and arginine diosynthesis on *Neurospora crassa. J. Bact.* **119**, 893-898.

Cariotis, M. &. Jones, R. F. (1974 b) Cross pathway regulation: Tryptophan mediated control of histidine, tryptophan and arginine biosynthesis in *Neurospora crassa. J. Bact.* **119**, 889-892.

Case, M. E. &. Giles, N. H. (1975) Genetic evidence on the organisation and action of the *qa-1* gene product: a protein regulating the induction of the three enzymes in quinate metabolism in *Neurospora crassa. Proc. Natl. Acad. Sci. USA* **72**, 553-557.

Case, M. E. &. Giles, N. H. (1976) Gene order in the *qa* gene cluster of *Neurospora crassa. Mol. Gen. Genet.* **147**, 83-89.

Chaleff, R. S. (1974 a) The inducible quinate-shikimate catabolic pathway in *Neurospora crassa*: Genetic organisation. *J. Gen. Microbiol.* **81**, 337-355.

Chaleff, R. S. (1974 b) The inducible quinate-shikimate catabolic pathway in *Neurospora crassa*: Induction and regulation of enzyme synthesis.
J. Gen. Microbiol. **81**, 357-372.

Charles, I. G., Keyte, J. W., Brammar, W. J. & Hawkins, A. R. (1985) Nucleotide sequence encoding the biosynthetic dehydroquinase function of the pentafunctional *arom* locus of *Aspergillus nidulans. Nucl. Acids Res.* **22**, 8119-8128.

Charles, I. G., Keyte, J. W., Brammar, W. J., Smith, M. & Hawkins, A. R. (1986) The isolation and nucleotide sequence of complete *arom* locus of *Aspergillus nidulans. Nucl. Acids Res.* **14**, 2201-2213.

Chaudhuri, S., Duncan, K., Graham, L. D. & Coggins, J. R. (1991) Identification of the active site lysine residues of two biosynthetic 3-dehydroquinases. *Biochem. J.* **275**, 1-6.

Cherry, J. R., Johnston, T. R., Dollard, C., Shuster, J. R. & Davis, C. L. (1989) Cyclic AMP-dependent protein kinase phosphorylates and inactivates the yeast transcriptional activator *ADR1*. *Cell* **56**, 409-419.

Choy, B. & Green, M. R. (1993) Eukaryotic activators function during multiple steps of pre-initiation complex assembly. *Nature* **366**, 531-536.

Clements, J. M. & Roberts, C. F. (1985) Molecular cloning of the 3-phosphoglycerate kinase (*PGK*) gene from *Aspergillus nidulans. Curr. Genet.* **9**, 293-298.

Clutterbuck, A. J. (1973) Gene symbols in *Aspergillus nidulans. Genet. Res.* **21**, 909-911.
Colgan, J., Wampler, S. & Manley, J. (1993) Interaction between a transcriptional activator and transcription factor II B *in vivo*. *Nature* **362**, 549-553.

Conaway, R. C. & Conaway, J. W. (1993) General initiation factors for RNA polymerase II. *Ann. Rev. Biochem.* **62**, 161-183.

Cornwell, E. V. & MacDonald, D. W. (1984) *glnA* mutations define a structural gene for glutamine synthetase in *Aspergillus*. *Curr. Genet.* **8**, 33-36.

Cote, J., Workman, J. L. & Peterson, C. L. (1994) Stimulation of *GAL4* derivative binding to nucleosome DNA by the yeast *SWI/SNF* complex. *Science* **265**, 53-60.

Cress, D. W. & Triezenberg, S. J. (1991) Critical structural elements of the VP16 transcriptional activation domain. *Science* **251**, 87-90.

Cubero, B. & Scazzochio, C. (1994) Two different , adjacent, and divergant zinc finger binding sites are necessary for *CREA* mediated carbon catabolite repression in the proline gene cluster of *Aspergillus nidulans*. *EMBO J.* **13**, 407-415.

Da Silva, A. J. S., Whittington, H. A. Clements, J. & Hawkins, A. R. (1986) Sequence analysis and transformation by the catabolic-3-dehydroquinase (*qutE*) gene from *Aspergillus nidulans*. *Biochem. J.* **240**, 481-486.

Davis, M. A., Kelly, J. M. & Hynes M. J. (1993) Fungal catabolic gene regulation; Molecular-Genetic analysis of the *amdS* gene of *Aspergillus nidulans*. *Genetica* **90**, 133-145.

Devereux, J., Haeberli, P. & Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.

Dingwall, C., Robbins, J., Dilworth, S. M., Roberts, B. & Richardson, W. D. (1988) The nucleoplasmin nuclear location sequence is larger and more complex than that of SV40 large-T antigen. *J. Cell. Biol.* **107**, 841-849.

Dingwall, C & Laskey, R. (1991) Nuclear targeting sequences-a consensus? *Trends Biochem. Sci.* **258**, 942-947.

Douglas, H. C. & Condie, F. (1954) The genetic control of galactose utilization in *Saccharomyces*. *J. Bacteriol.* **68**, 662-670.

Douglas, H. C. & Pelroy, G. (1963) A gene controlling inducibility of the galactose pathway enzymes in *Saccharomyces*. *Biochim. Biophys. Acta* **65**, 155-156.

Douglas, H. C. &. Hawthorne, D. C. (1966) Regulation of genes controlling synthesis of the galactose pathway enzymes in yeast. *Genetics* **54**, 911-916.

Dowzer, C. E. A. &. Kelly, J. M. (1989) Cloning the *creA* gene from *Aspergillus nidulans*: a gene involved in carbon catabolite repression. *Curr. Genet.* **15**, 457-459.

Dowzer, C. E. A. &. Kelly J. M. (1991) Analysis of the *creA* gene a regulator of carbon catabolite repression in *Aspergillus nidulans*. *Mol. Cell. Biol.* **11**, 5701-5709.

Dugaiczyk, A., Boyer, H. W. & Goodman, H. M. (1975) Ligation of *Eco*RI endonuclease generated DNA fragments in linear and circular structures. *J. Mol. Biol.* **96**, 171-184.

Duncan, K., Edwards, R. M. & Coggins, J. R. (1987) The pentafunctional *AROM* enzyme of *Saccharomyces cerevisiae* is a mosaic of monofunctional domains. *Biochem. J.* **246**, 375-386.

Ehinger, A., Denison, S. H. & May, G. S. (1990) Sequence organisation and expression of the core histone genes of *Aspergillus nidulans*. *Mol. Gen. Genet.* **222**, 416-424.

Enos, A. P. & Morris, N. R. (1990) Mutation of a gene that encodes a kinesin like protein blocks nuclear division in *Aspergillus nidulans*. *Cell* **60**, 1019-1027.

Espeso, E. A. & Penalva, M. A. (1994) *In-vitro* binding of the two finger repressor *CREA* to several concensus and non-concensus sites in the *ipnA* upstream region is context dependent. *FEBS Letts* **342**, 43-48.

Fantes, P. A. & Roberts, C. F. (1973) β-galactosidase activity and lactose utilisation in *Aspergillus nidulans*. *J. Gen. Microbiol.* **77**, 471-486.

Felenbok, B., Sequeval, D., Mathieu, M. Sibley, S., Gwynne, D. I. & Davies, R. W. (1988) The ethanol regulon in *Aspergillus nidulans* characterisation and sequence of the positive regulatory gene *alcR*. *Gene* **73**, 385-396.

Feldherr, C. M., Kallenbach, E. & Shultz, N. (1984) Movement of karyophilic protein through the nuclear pores of oocytes. *J. Cell Biol.* **99**, 2216-2222.

Fridden, P. & Schimel, P. (1988) *LEU3* of *Saccharomyces cerevisiae* activates multiple genes for branched-chain amino acid biosynthesis by binding to a common decanucleotide core sequence. *Mol. Cell. Biol.* **8**, 2690-2697.

Furst, P., Hu, S., Hackett, R. & Hammer, D. (1988) Copper activates metallothionine gene transcription by altering the conformation of a specific DNA binding protein. *Cell.* **55**, 705-717.

Gaertner, F. H., Ericson, M. C. & Coggins, J. R. (1970) The pentafunctional *arom* enzyme of *Saccharomyces cerevisiae* is a mosaic of monofunctional domains. *Biochem. J.* **246**, 375-386.

Garbe, T., Servos, S., Hawkins, A. R., Dimitriadis, G., Young, G., Dougan, G. & Charles, I.G. (1991) The *mycobacterium tuberculosis* shikimate pathway genes: Evolutionary relationship between biosynthetic and catabolic 3-dehydroquinases. *Mol. Gen. Genet.* **228**, 385-392.

Geever, R. F., Huiett, L., Baum, J. A., Tyler, B. M., Patel, V. B., Putledge, B. J., Case, M. E. & Giles, N. H. (1989) DNA sequence organisation and regulation of the *qa* gene cluster of *Neurospora crassa*. *J. Mol. Biol.* **207**, 15-34.

Geever, R. F. (1991) Aspergillus nidulans negative acting regulatory protein (*qutR*) gene, complete cds, and adjacent unidentified gene. *EMBL database*. Acession number **M59936**.

Giles, N. H., Case, M. E., Partridge, C. W. H. & Ahmed, S. I. (1967) A gene cluster coding for an aggregate of five aromatic synthetic enzymes. *Proc. Natl. Acad. Sci. USA* **58**, 1453-1460.
Giles, N. H. (1978) The organisation, function and evolution of gene clusters in eucaryotes. *Nature* **112**, 641-657.

Giniger, E., Varnum, S. M. & Ptashne, M. (1985) Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* **40**, 767-774.

Gorfinkiel, L., Dallinas, G. & Scazzochio, C. (1993) Sequence and regulation of the *uapA* gene encoding uric acid xanthine permease in the fungus *Aspergillus nidulans*. *J. Biol. Chem.* **268**, 23376-23381.

Grant, S., Roberts, C. F., Lamb, H., Stout, M. & Hawkins, A. R. (1988) Genetic regulation of the quinic acid utilization (*qut*) gene cluster in *Aspergillus nidulans*. *J. Gen. Microbiol.* **134**, 347-358.

Gyllensten, U. B. &. Erlich, H. A. (1988) Generation of single stranded DNA by the polymerase chain reaction and its application to direct sequencing of the *HLA-DQA* locus. *Proc. Natl. Acad. Sci. USA* **85**, 7652-7656.

Gurr, S. J., Unkles, S. E. & Kinghorn, J. R. (1987) The structure and organisation of nuclear genes of filamentous fungi. *in Kinghorn, J. R. (Ed), Gene structure in Eukaryotic Microbes. Soc. Gen. Microbiol. Spec. Pub. IRL press, Oxford*, 93-139.

Hall, M. N., Hereford, L. & Herskowitz, I. (1984) targeting of *E.coli* β-galactosidase to the nucleus in yeast. *Cell* **36**, 1057-1065.

Halvorsen, Y. P. C., Nandabalan, K. & Dickson, R. C. (1991) Identification of base and backbone contacts used for DNA-sequence recognition and high -affinity binding by *lac9*, a transcription activator containing a C_6 zinc finger. *Mol. Cell. Biol.* 11, 1777-1784.

Hawkins, A. R., Giles, N. H. & Kinghorn, J. R. (1982) General and biochemical aspects of quinate breakdown in the filamentous fungus *Aspergillus nidulans*. *Biochem. Genet.* **20**, 271-286.

Hawkins, A. R., Da Silva, A. J. S. & Roberts, C.F. (1984) Evidence for two control genes regulating expression of the quinic acid (*qut*) utilisation cluster in *Aspergillus nidulans*. *J. Gen. Microbiol* **130**, 567-574.

Hawkins, A. R., Da Silva, A. J. S. & Roberts, C.F. (1985) Cloning and characterisation of the three enzme structural genes *qutB*, *qutC* and *qutE* from the quinic acid utilisation gene cluster in *Aspergillus nidulans*. *Curr. Genet.* **9**, 305 311.

Hawkins, A. R. (1987) The complex *arom* locus of *Aspergillus nidulans*. *Curr. Genet.* **11**, 491-498.

Hawkins, A. R., Lamb, H. K., Smith, M., Keyte, J. W. & Roberts, C.F. (1988) Molecular organisation in quinic acid (*qut*) utilisation gene cluster in *Aspergillus nidulans*. *Mol. Gen. Genet.* **214**, 224-231.

Hawkins, A. R., Gurr, S. J., Montague, P. & Kinghorn, J. R. (1989) Nucleotide sequence and regulation of expression of the *Aspergillus nidulans gdhA* gene encoding NADP-dependant Glutamate dehydrogenase. *Mol. Gen. Genet.* **218**, 105-111.

Hawkins, A. R. & Smith, M. (1991) Domain structure and interaction within the pentafunctional *arom* polypeptide. *Eur. J. Biochem.* **196**, 717-724.

Hawkins, A. R., Lamb, H. K. & Roberts, C.F. (1992) Structure of the *Aspergillus nidulans qut* repressor-encoding gene; implications for the regulation of transcription initiation. *Gene* **110**, 109-114.

Hawkins, A. R., Lamb, H. K. & Roberts, C.F. (1993) Genesis of eukaryotic transcriptional activator and repressor proteins by splitting a multidomain anabolic enzyme. *Gene* **136**, 49-54.

Hawkins, A. R., Lamb, H. K. & Moore, D. M. (1994) Evolution of transcription regulatory proteins by enzyme recruitment: molecular models for nitrogen metabolite repression and ethanol utilisation in eukaryotes. *Gene* **146**, 145-158.

Hinnebusch, A. G. (1985) A hierarchy of trans-acting factors modulates translation of an activator in the general amino acid biosynthetic genes in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **5**, 2349-2360.

Hinnebusch, A. G. &. Fink, G. R. (1983) Positive regulation in the general control of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **80**, 5347-5378.

Hinnebusch, A. G. (1988) Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. *Microbiol. Rev.* **52**, 248-273.

Hope, I. A., Mahadevan, S. & Struhl, K. (1988) Structural and functional characterisation of the short acidic transcriptional activation region of yeast GCN4 protein. *Nature* **333**, 635-640.

Hope, I. A. &. Struhl, K. (1985) *GCN4* protein, synthesized *in vitro*, binds *HIS3* regulatory sequences: Implications for general control of amino acid biosynthetic genes in yeast. *Cell* **43**, 177-188.

Hope, I. A. &. Struhl, K. (1986) Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**, 885-894.

Hopper, J. E. & Rowe, L. B. (1978) Molecular expression and regulation of the galactose pathway genes in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **253**, 7566-7569.

Huiet, L. (1984) Molecular analysis of the *Neurospora qa-1* regulatory region indictes that two interacting genes control *qa* gene expression. *Proc. Natl. Acad. Sci. USA* **81**, 1174-1178.

Hynes, M. J. (1974) Effect of ammonium, L-glutamate and L-glutamine on nitrogen catabolism in *Aspergillus nidulans*. *J. Bact.* **120**, 1116-1123

Hynes, M. J. (1975) Studies of the role of the *areA* gene in the regulation of nitrogen metabolism in *Aspergillus nidulans*. *Aust. J. Biol. Sci* **28**, 301-313.

Hynes, M. J. (1977) Induction of the acetamidase gene of *Aspergillus nidulans* by acetate metabolism. *J. Bacteriol.* **131**, 770-775.

Hynes, M. J. and Kelly, J. M. (1977) Plieotropic mutants of *Aspergillus nidulans* altered in carbon catabolism. *Mol. Gen. Genet.* **150**, 193-204.

Hynes, M. J., Corrick, C. M., Kelly, J. M. & Littlejohn, T. G. (1988) Identification of the sites of action of regulatory genes controlling the *amdS* gene of *Aspergillus nidulans*. *Mol. Cell. Biol.* **8**, 2589-2596.

Hynes, M. J., Andrianopoulos, A., Davis, M. A., Van Heeswijck, R., Katz, M. E., Littlejohn, T. G., Richardson, I. B. & Saleeb, J. A. (1989) Multiple circuits regulating the *amdS* gene of *Aspergillus nidulans*. *in Molecular Biology of Filamentous Fungi. eds Nevalainen, H. & Penttila, M.* pp63-72.

Iwahana, H., Yoshimoto, K., & Itakura, M., (1992) Detection of point mutations by SSCP of PCR amplified DNA after endonuclease digestion. *Biotechniques*. **12**, 64-66.

Jacobson, J. W., Hart, B. A., Doy, C. H. & Giles, N. H. (1972) Purification and stability of the multienzyme complex encoded in the *arom* gene cluster of *Neurospora crassa*. *Biochim. Biophys. Acta.* **289**, 1-12.

Jensen, R. A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol* **30**, 409-425.

Johnston, M., Flick, J. S. & Pexton, T. (1994) Multiple mechanisms provide rapid and stringent glucose repression of *GAL* gene expression in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **14**, 3834-3841.

Johnston, M. &. Davis, R. (1984) Sequences that regulate the divergant *GAL1-GAL10* promoter in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **4**, 1440-1448.

Kalderon, D., Richardson, W.D., Markham, A. F. & Smith, A.E. (1984) Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* **311**, 33-38.

Kammerer, B., Guyonvarch, A. & Hubert, J. C. (1984) Yeast regulatory gene *PPR1*: nucleotide sequence, restriction map and codon usage. *J. Mol. Biol.* **180**, 239-250.

Katz, M. E. & Hynes, M. J. (1989) Characterisation of the *amdR* controlled *lamA* and *lamB* genes of *Aspergillus nidulans*. *Genetics* **122**, 331-339.

Katz, M. E. & Hynes, M. J. (1989) Isolation and characterisation of the acetate regulatory gene *facB* from *Aspergillus nidulans*. *Mol. Cell. Biol.* **9**, 5696-5701.

Kinghorn, J. R. & Pateman, J. A. (1973) NAD and NADP L-glutamate dehydrogenase activity and ammonium regulation in *Aspergillus nidulans*. *J. Gen. Microbiol.* **78**, 39-46.

Kinghorn, J. R. & Pateman, J. A. (1975) Studies of partially derepressed mutants at the tamA and *areA* loci in *Aspergillus nidulans*. *Mol. Gen. Genet.* **140**, 137-147.
Kinghorn, J. R. & Hawkins, A. R. (1982) Genetical and biochemical aspects of quinate breakdown in the filamentous fungi *Aspergillus nidulans*. *Mol. Gen. Genet.* **186**, 145-152.

Kleanthous, C., Deka, R., Davis, K., Kely, S. M., Cooper, A., Harding, S. E., Price, N. C., Hawkins, A. R. & Coggins, J. R. (1992) A comparison of the enzymological and biophysical properties of two distinct classes of dehydroquinase enzymes. *Biochem. J.* **282**, 687-695.

Krulis, P. J., Raine, A. R. C., Gadhavi, P. L. & Leiue, E. D. (1992) Structure of the DNA-binding domain of zinc GAL4. *Nature* **356**, 448-450.

Kudla, B., Caddick, M. X., Langden, T., Martinez-Rossi, N. M., Bennet, C. F., Sibley, S., Davies, R. W. & Arst, N. H. (1990) The regulatroy gene *areA* mediating nitrogen metabolite repression in *Aspergillus nidulans*. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. *EMBO J.* **9**, 1355-1364.

Kulmberg, P., Prange, T., Mathieu, M., Sequeval, D., Scazzocchio, C. & Felenbok, B. (1991) Correct intron splicing generates a new type of a putative zinc binding domain in a transcriptional activator of *Aspergillus nidulans*. *FEBS Lett.* **280**, 11-16.

Kulmberg, P., Sequeval, D., Leuouval, F., Mathieu, M. & Felenbok, B. (1992) Identification of a promoter region involved in the autoregulation of the transcriptional activator ALCR in *Aspergillus nidulans*. *Mol. Cell. Biol.* **12**, 1932-1939.

Kuswandi, & Roberts, C. F. (1992) Genetic control of the protocatechuic acid pathway in *Aspergillus nidulans*. *J. Gen. Micro.* **138**, 817-823.

Lamb, H. K., Hawkins, A. R., Smith, M., Harvey,I. J., Brown, J., Turner, G. & Roberts, C.F. (1990) Spatial and biological characterisation of the complete quinic acid utilisation cluster in *Aspergillus nidulans*. *Mol. Gen. Genet.* **223**, 17-23.

Lamb, H. K., Bagshaw, C. R. & Hawkins, A. R. (1991) *In vivo* overproduction of the pentafunctional AROM polypeptide in *Aspergillus nidulans* affects metabolic flux in the quinate pathway. *Mol. Gen. Genet.* **227**, 187-196.

Lamb, H. K., Hawkins, A. R. & Roberts, C. F. (1992) A second gene (*qutH*) within the *Aspergillus nidulans* quinic acid utilisation gene cluster encodes a protein with a putative zinc cluster motif. *Gene* **112**, 219-224.

Lambert, J. M., Boocock, M. R. & Coggins, J. R. (1985) The 3-dehydroquinate synthase activity of the pentafunctional arom complex of *Neurospora crassa* is Zn^{2+} dependant. *Biochem. J.* **226**, 817-829.

Lamphier, M. S. & Ptashne, M. (1992) Multiple mechanisms mediate glucose repression of the yeast *GAL1* gene. *Proc. Natl. Acad. Sci. USA* **89**, 5922-5926.

Laughon, A. &. Gesteland, R. F. (1984) Primary structure of the Saccharomyces cerevisiae *GAL4* gene. *Mol. Cell. Biol.* **4**, 260-267.

Laybourn, P. J. &. Dahmous, M. E. (1990) Phosphorylation of RNA polymerase II occurs subsequennt to interaction with the promoter and before the initiation of transcription. *J. Biol. Chem.* **265**, 13165-13173.

Leff, S. E., Rosenfeld, M. G. & Evans, R. M. (1986) Complex transcriptional units: diversity in gene expression by Alternative RNA processing. *Ann. Rev. Biochem.* **55**, 1091-1117.

Leuther, K. K., Salmeron, J. M. & Johnston, S. A. (1993) Genetic evidence that an activation domain of GAL4 does not require acidity and may form a β-sheet. *Cell* **72**, 575-585.

Leuther, K. K. &. Johnson, S. A. (1991) Non-dissociation of GAL4 and GAL80 *in vivo* after galactose induction. *Science* **256**, 1333-1335.

Levesley, I., Newton, G. H., Lamb, H.K., van Schothorst, E., Dalgleish, R. W. M., Samson, A. C. R., Roberts, C. F. & Hawkins, A. R. (1996) Domain structure and function within the QUTA protein *of Aspergillus nidulans*: implications for the control of transcription. *Microbiology* **142**, 87-98.

Lockington, R. A., Scazzochio, C., Sequeval, D., Mathieu, M. & Felenbok, B. (1987) Regulation of *alcR*, the positive regulatory gene of the ethanol utilisation regulon of *Aspergillus nidulans*. *Mol. Microbiol.* **1**, 275-281.

Loo, S. & Rine, J. (1994) Silencers and domains of generalised repression. *Science* **264**, 1768-1771.

Lowe, T., Sharefkin, J., Yang, S. Q. & Dieffenbach, W. (1990) A computer program for selection of oligonucleotide primers for polymerase chain reactions. *Nucl. Acids. Res.* **18**, 1757-1761.

Lumbsden, J. & Coggins, J. R. (1977) The subunit structure of the AROM multienzyme complex of *Neurospora crassa*. *Biochem. J.* **161**, 599-607.

Lumbsden, J. &. Coggins, J. R. (1978) The subunit structure of the AROM multienzyme. *Biochem. J.* **169**, 441-444.

Ma, J. & Ptashne, M. (1987a) Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* **48**, 847-853.

Ma, J. & Ptashne, M. (1987b) The carboxy terminal 30 amino acids of GAL4 are recognised by GAL80. *Cell* **50**, 137-142.

Ma, J. & Ptashne, M. (1987c) A new class of yeast transcriptional activators. *Cell* **51**, 113-119.

MacCabe, A. P., Vanliempt, H., Palissa, H., Unkles, S. E., Riach, M. B. R., Pfiefer, E., Vondohren, H. & Kinghorn, J. R. (1991) Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from *Aspergillus nidulans*-molecular characterization of the *acvA* gene encoding the first enzyme of the penicillin biosynthetic pathway. *J. Biol. Chem.* **266**, 12646-12654.

Mackay, K., Lund, A. M., Ragunath, M., Steinmann, B. & Dalgleish, R. (1993) SSCP detection of a Gly 565 Val substitution in Pro-α1 (I) collagen chain resulting in osteogenesis imperfecta type II. *Human Genet.* **91**, 439-444.

Malim, M. H., Hauber, J., Le, S.-Y., Maizel, J. V. & Cullen B. R. (1989) The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature.* **338**, 254-257.

Mandel, M. & Higa, A. (1970) Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **53**, 159-162.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular cloning (a laboratory manual). *Cold Spring Harbor Laboratory Press. Cold Spring Harbor, N. Y.*

Marczak, J. E. & Brandriss, M. C. (1989) Isolation of constitutive mutations affecting the proline utilisation pathway in *Saccharomyces cerevisiae* and molecular analysis of the *PUT3* transcriptional activator. *Mol. Cell Biol.* **9**, 4696-4705.

Marczak, J. E. & Brandriss, M. C. (1991) Analysis of constitutive and non-inducible mutants of the *PUT3* transcriptional activator. *Mol. Cell Biol.* **11**, 2609-2619.

Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. (1992) DNA recognition by GAL4 structure of a protein DNA complex. *Nature* **356**, 408-414.

Marmorstein, R. & Harrison, S. C. (1994) Crystal-structure of a *PPR1*-DNA complex, DNA recognition by proteins containing a $Zn_{(2)}$ $Cys_{(6)}$ bi-nuclear cluster. *Genes Dev.* **8**, 2504-2512.

Marshall, M. A. & Timberlake, W. E. (1990) *Aspergillus nidulans wetA* activates spore specific gene expression. *Mol. Cell. Biol.* **11**, 55-62.

Marzluff, W. F. & Pandey, N. B. (1988) Multiple regulatory steps control histone mRNA concentration. *Trends in Biochemical Sciences.* **13**, 49-52.

Matern, M. & Holzer, H. (1977) Catabolite inactivation of the galactose uptake system in yeast. *J. Biol. Chem.* **252**, 6399-6402.

Mathieu, M. &. Felenbok, B. (1994) The *Aspergillus nidulans CREA* protein mediates glucose repression of the ethanol regulon at various levels through competition with the *alcR* specific transactivator. *EMBO J.* **13**, 4022-4027.

May, G. S. & Morris, N. R. (1986) Developmental regulation of the *tubC* beta-tubulin gene of *Aspergillus nidulans. J. Cell. Biol.* **103**, A144. Messing, J., Crea, R. & Seeburg, P. H. (1981) A system for shotgun DNA sequencing. *Nucleic Acids Res.* **9**, 309-321.

Melchior, F. & Gerace, L. (1995) Mechanisms of nuclear protein import. *Curr. Opin. Cell Biol.* **7**, 310-318.

Michaud, J., Brody, L. C., Steel, G., Fontaine, G., Martin, L. S., Valle, D. & Mitchell, G. (1992) Strand-separating conformational polymorphism analysis: Efficacy of detection of point mutations in the human ornithine d-aminotransferase gene. *Genomics.* 13, 389-394.

Miller, J. H. (1972) *in Experiments in Molecular Genetics. ed J. H. Miller. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, N. Y.*

Mitchell, P. J. & Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence specific DNA-binding proteins. *Science* **245**, 371-378.

Montenegro, E., Fierro, F., Fernandez, F. J., Gutierrez, S. & Martin, J. F. (1992) Resolution of chromosome III and chromosome VI of *Aspergillus nidulans* by pulse-field gel electrophoresis shows that the penicillin biosynthetic pathway genes, gene *pcbAB*, gene *pcbC* and gene *penDE* are clustered on chromosome VI (3.0 megabases). *J. Bact.* **174**, 7063-7067.

Moore , J. D. & Hawkins, A. R. (1993) Overproduction of, and interaction within, bi-functional domains from the amino and C termini of the pentafunctional AROM protein of *Aspergillus nidulans. Mol. Gen. Genet.* **240**, 92-102.

Moore, J. D., Coggins, J. R., Virden, R. & Hawkins, A. R. (1994) Efficient independent activity of a monomeric monofunctional dehydroquinase synthase derived from the N-terminus of the pentafunctional AROM protein of *Aspergillus nidulans. Biochem. J.* **301**, 297-304.

Mueller, P. P. & Hinnebusch, A. G. (1986) Multiple upstream AUG codons mediate translational control of *GCN4. Cell* **45**, 201-207.

Mueller, P. P., Harashima, S. & Hinnebusch, A. G. (1987) A segment of *GCN4* mRNA containing the upstream AUG codons confers translational control upon a heterologous yeast transcript. *Proc. Natl. Acad. Sci. USA* **84**, 2863-2867.

Mylin, L. M., Johnston, M. & Hopper, J. E. (1990) Phosphorylated forms of GAL4 are correlated with ability to activate transcription. *Mol. Cell. Biol.* **10**, 4623-4629.

Neuwald, N. F., York, J. D. & Majerus, P. W. (1991) Diverse proteins homologous to inositol phosphatase. *FEBS Lett.* **229**, 16-18.

Niederberger, P., Miozzari, G. & Hutter, R. (1981) Biological role of the General Control of amino-acid biosynthesis in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **1**, 584-593.

Nogi, Y., & Fukasawa, T. (1984) Nucleotide sequence of the yeast regulatory gene *GAL80. Nucl. Acids. Res.* **12**, 9287-9298.

O'Brien, T., Hardin, S., Greenleaf, A. & Lis, J. T. (1994) Phosphorylation of RNA polymerase II C terminal domain and transcription elongation. *Nature* **370**, 75-77.

O'Connell, M. J., Osmani, A. H., Morris, N. R. and Osmani, S. A. (1992) An extra copy of *nimE* cyclin B elevates preMPF levels and partially suppresses mutations of *nimT cdc25* in *Aspergillus nidulans*. *EMBO J.* **11**, 2139-2149.

Oakley, B. R., Oakley, C. E., Kniepkamp, K. S. and Rinehart, J. E. (1985) Isolation and characterization of cold sensitive mutations at the *benA* beta tubulin locus of *Aspergillus nidulans*. *Mol. Gen. Genet.* **201**, 56-64.

Ono, Y., Onada, H., Sasada, R., Igarashi, K., Sugino, Y. & Nishioka, K. (1983) The complete nucleotide sequence of the cloned hepatitis B virus DNA; subtype adr and adw. *Nucl. Acids Res.* **11**, 1747-1757.

Orita, M., Iawahana, H., Kanazawa, H., Hayaashi, K. & Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* **86**, 2766-2770.

Pan, T. &. Coleman, E. (1990) GAL4 transcription factor is not a "zinc finger" but forms a Zn(II)2 Cys6 binucleur cluster. *Proc. Natl. Acad. Sci. USA* **87**, 2077-2081.

Partridge, C. W., Case, M. E. & Giles, N. H. (1972) Direct induction of wild-type *Neurospora crassa* of mutants (*qa-1c*) constitutive for the catabolism of quinate and shikimate. *Genetics* **72**, 411-417.

Patel, V. B., Schweizer, M., Dykstra, C. C., Kushner, S. R. & Giles, N. H. (1981) Genetic organization and transcription regulation in the *qa* gene cluster of *Neurospora crassa*. *Proc. Natl. Acad. Sci. USA* **78**, 5783-5787.

Pateman, J. A., Kinghorn, J. R., Dunn, E. & Forbes, E. (1973) Ammonium regulation in *Aspergillus nidulans*. *J. Bact.* **114**, 943-950.

Pateman, J. A., Doy, C. H., Olsen, J. E., Norris, U., Creaser, E. H. & Hynes, M. (1983) Regulation of alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (AldDH) in *Aspergillus nidulans*. *Proc. Royal. Soc. London* Series B. **217**, 243-264.

Payne, J. M., Laybourn, P. J. & Dahmous, M. E. (1989) The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl terminal domain of subunit IIa. *J. Biol. Chem.* **264**, 19621-19629.

Penn, M. D., Golgoci, B. & Greer, H. (1983) Identification of *AAS* genes and their regulatory control of Amino Acid biosynthesis in yeast. *Proc. Natl. Acad. Sci. USA* **80**, 2704-2708.

Pfeifer, K. Kim, K. S., Kogan, S. & Guarente, L. (1989) Functional dissection and sequence of yeast *HAP1* activator. *Cell* **56**, 291-301.

Piotrowaska, M. (1980) Cross pathway regulation of orthinine carbomyltransferase synthsis in *Aspergillus nidulans*. *J. Gen. Microbiol.* **116**, 335-339.

Polkinghorne, M. &. Hynes, M. J. (1975) Mutants affecting histidine utilisation in *Aspergillus nidulans*. *Genet. Res.* **25**, 119-135.

Pontecorvo, G., Roper., J. A., Hemmons, L. M., MacDonald, K. D. & Bufton, A. W. J. (1953) The genetics of *Aspergillus nidulans*. *Adv. Genet.* **5**, 141-238.

Powers, M. A. & Forbes, D. J. (1994) Cytoplasmic factors in nuclear transport: what's importin? *Cell* **79**, 931-934.

Prade, R. A. &. Timberlake, W. E. (1993) The *Aspergillus nidulans brlA* regulatory locus consists of overlapping transcription units that are individually required for conidiophore development. *EMBO J.* **12**, 2439-2447.

Ptashne, M. (1988) How transcriptional activators work. *Nature* **335**, 683-689.

Punt, P. J., Dingamanse, M. A., Kuyvehoven, A., Soeda, R. D. M., Pouwels, P. H. & den Hondel, C. A. M. J. J. (1990) Functional elements in the promoter region of the *Aspergillus nidulans gpdA* gene encoding glceraldehyde-3-phosphate dehydrogenase. *Gene* **93**, 101-109.

Raghow, R. (1987) Regulation of messenger RNA turnover in eukaryotes. *Trends in Biochemical Sciences*. **12**, 358-360.
Reece, R. J. & Ptashne, M. (1993) Determinants of binding site specificity among C6 zinc cluster proteins. *Science* **261**, 909 911.

Regier, J. L., Shen, F., Triezenberg, S. J. (1993) Pattern of aromatic and hydrophobic amino acids critical for one of the 2 subdomains of the VP16 transcriptional activator. *Proc. Natl. Acad. Sci. USA* **90**, 883-887.

Reichelt, R., Holzenburg, A., Buhle, E. L., Jarnik, M., Engel, A. & Aebi, U. (1990) Correlation between structure and mass distribution of the nuclear pore complex, and of distinct pore complex components. *J. Cell Biol.* **110**, 883-894.

Remboutsika, E. & Kohlhaw G. B. (1994) Molecular architecture of a Leu3p-DNA complex in solution: a biochemical approach. *Mol. Cell Biol.* **14**, 5547-5557.

Richardson, I. B., Katz, M. E. & Hynes, M. J. (1992) Molecular characterization of the *lam* locus and sequences involved in regulation by the AMDR protein of *Aspergillus nidulans*. *Mol. Cell. Biol.* **12**, 337-364.

Roberts, C. F. (1963) The genetics of carbohydrate utilization in *Aspergillus nidulans*. *J. Gen. Microbiol.* **31**, 45-48.

Roberts, C. F. (1969) Silica gel stock cultures of *Aspergillus nidulans*. *Aspergillus News Letters* **10**, 29.

Roper, J. A. (1952) Production of heterozygous diploids in filamentous fungi. *Experientia. Basel.* **8**, 14-15.

Rotman, B. & Spiegleman, S. (1953) The conversion of negatives to positives in 'slow' adapting populations of yeast. *J. Bacteriol.* **66**, 492-497.

Roussou, I., Thireos, G. & Hauge, B. M. (1988) Transcriptional-translational regulatory circuits in *Saccharomyces cerevisiae* which involves the *GCN4* transcriptional activator and the *GCN2* protein kinase. *Mol. Cell. Biol.* **8**, 2132-2139.

Sadowski, I., Niedbala, D., Wood, K. & Ptashne, M. (1991) GAL4 is phosphorylated as a consequence of transcriptional activation. *Proc. Natl. Acad. Sci. USA* **88**, 10510-10514.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1988) Enzymatic amplification of β-globin sequences and restriction analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-1354.

Saleeba, J. A., Cobbett, C. S. & Hynes, M. J. (1992) Characterization of the *amdA* regulated *aciA* gene of *Aspergillus nidulans*. *Mol. Gen. Genet.* **235**, 349-358.

Salmeron, J. M. &. Johnston, S. A. (1986) Analysis of the *Kluyveromyces lactis* positive regulatory gene *LAC9* reveals functional homology to but sequence divergence from, the *Saccharomyces cerevisiae GAL4* gene. *Nucl. Acids Res.* **14**, 7767-7781.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977) DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Sarkar, G., Yoon, H. S. & Sommer, S. S. (1992) Screening for mutations by RNA single-stranded conformation polymorphism (rSSCP); comparison with DNA-SSCP. *Nucl. Acids. Res.* **20**, 871-878.

Savov, A., Angelicheva, D., Jordanova, A., Eigel, A. & Kalagdjieva, L. (1992) High percentage acrylamide gels improve resolution in SSCP analysis. *Nucl. Acids Res.* **20**, 6741-6742.

Sequeval, D. &. Felenbok, B. (1994) Relationship between zinc content and DNA binding activity of the DNA binding motif of the transcription factor ALCR in *Aspergillus nidulans*. *Mol. Gen. Genet.* **242**, 33-39.

Sheffield, V. C., Beck, J. S., Kwitek, A. E., Sandstrom, D. W. & Stone, E. M. (1993) The sensitivity of single-strand conformational polymorphism analysis for the detection of single base substitutions. *Genomics* **16**, 325-332.

Shroff, R.A., Lockington, R. A. & Kelly, M. J. (1996) Analysis of mutations in the *creA* gene involved in carbon catabolite repression in *Aspergillus nidulans*. *Canadian J. Micro.* **42**, 950-959.

Siddqui, A. H. & Brandriss, M. C. (1989) The *Saccharomyces cerevisiae PUT3* activator associates with proline-specific upstream activator sequences. *Mol. Cell Biol.* **9**, 4705-4712.

Silver, P. A., Keegan, L. P. & Ptashne, M. (1984) Amino terminus of the yeast *GAL4* gene product is sufficient for nuclear localization. *Proc. Natl. Acad. Sci. USA* **81**, 5951-5955.

Silver, P. A., Brent, R. & Ptashne, M. (1986) DNA binding is not sufficient for nuclear localization of regulatory proteins in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **6**, 4763-4766.

Silver, P. A. (1986) Nuclear protein localization in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **33**, 415-420.

Siomi, H. & Dreyfuss, G. (1995) A nuclear localization domain in the hnRNP A1 protein. *J. cell Biol.* **129**, 551-560.

Speigleman, S., Rotman-Sussman, R. & Pinska, E. (1950) The cytoplasmic nature of 'long term adaptation' in yeast. *Proc. Natl. Acad. Sci. USA* **36**, 591-606.

Spinardi, L., Mazars, R. & Thiellet, C. (1991) Protocols for an improved detection of point mutations by SSCP. *Nucl Acids Res.* **19**, 4009.

Stallings, W. C., Abdel-Meguid, S. S., Lim, L. W., Shieh, H. S., Dayringer, H. E., Leimgruber, N. K., Stegeman, R. A., Anderson, K. S., Sikorski, J. A., Padgette, S. R. & Kishore, G. M. (1991) Structure and topological symmetry of the glyphosphate target 5-enol-pyruvylshikimate-3-phosphate synthase: A distinctive protein fold. *Proc. Natl. Acad. Sci. USA* **88**, 5046-5050.

Struhl, K. (1989) Molecular mechanisms of transcriptional regulation in yeast. *Annu. Rev. Biochem.* **58**, 1051-1077.

Sze, J. Y., Woontner, M., Jaehning, J. & Kohlhaw, G. B. (1992) *In vitro* transcriptional activation by a metabolic intermediate: activation of *LEU3* depends on α-isopropylmalate. *Science* **258**, 1143-1145.

Sze, J. Y., Remboutsika, E. and Kohlhaw, G. B. (1993) Transcriptional regulator *LEU3* of *Saccharomyces cerevisiae* separation of activator and repressor functions. *Mol. Cell. Biol.* **13**, 5702-5709.

Taylor, W. E. & Young, E. T. (1990) cAMP dependant phosphorylation and inactivation of yeast transcriptional factor does not affect DNA binding. *Proc. Natl. Acad. Sci. USA* **87**, 4098-4102.

Tobin, M. B., Fleming, M. D., Skatrud, P. L. & Miller, J. R. (1990) Molecular characterization of the acyl coenzymeA Isopenicillin-N acetyltransferase gene *penDE* from *Penicillium chrysogenum* and *Aspergillus nidulans* and activity of a recombinant enzyme in *Escherichia coli*. *J. Bact.* **172**, 5908-5914.

Torchia, T. E. & Hopper, J. E. (1986) Genetic and molecular analysis of the *GAL3* gene in the expression of the galactose/melibiose regulon of *Saccharomyces cerevisiae*. *Genetics* **113**, 229-246.

Tzamarias, D. & Thireos, G. (1988) Evidence that the GCN2 protein kinase regulates reinitiation by yeast ribosomes. *EMBO J.* 7, 3547-3551.

Unkles, S. E., Hawker, K. L., Grieve, C., Campbell, E. I. Montague, P. & Kinghorn, J. R. (1991) *crnA* encodes a nitrate transporter in *Aspergillus nidulans*. *Proc. Natl. Acad. Sci. USA* **88**, 204-208.

Vallee, B. L., Coleman, J. E. & Auld, D. S. (1991) Zinc fingers, Zinc clusters and Zinc twists in DNA binding protein domains. *Proc. Natl. Acad. Sci. USA* **88**, 999-1003.

van den Brink, J., van Zeijl, C., van den Hondel, C., & van Gorcom, R. (1990) Cloning and characterization of the NADPH cytochrome P450 oxidoreductase (*cprA)* gene of *Aspergillus niger*. *EMBL database* Z26938.

Van Hoy, M., Leuther, K. K., Kodadek, T. & Johnston S. A. (1993) The acidic activation domains of the GCN4 and GAL4 proteins are not α-helical but form β-sheets. *Cell* **72**, 587-594.

Vogelstein,B. & Gillespie, D. (1979) Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. USA* **76**, 615-619.

Wagner, P. & Hall, M. N. (1993) Nuclear protein transport is functionaly conserved between yeast and higher eukaryotes. *FEBS Lett.* **321**, 261-266.

Walker, M. S. & De Moss, J. A. (1983) Purification and characterization of the trifunctional β-subunit of anthranilite synthase from *Neurospora crassa*. *J. Biol. Chem.* **258**, 3571-3575.

Walker, M. S. & De Moss, J. A. (1986) Organisation of the functional domains of anthranilite synthase from *Neurospora crassa*. Limited proteolysis studies. *J. Biol. Chem.* **261**, 16073-16077.

Wek, R. C., Jackson, B. M. & Hinnebusch (1989) Juxtaposition of domains homologous to protein kinases and histidyl tRNA synthetases in GCN2 protein suggests a mechanism for coupling *GCN4* expression to amino acid availability. *Proc. Natl. Acad. Sci. USA* **86**, 4579-4583.

Whittington, H. A., Grant, S. Roberts, C. F., Lamb, H. K. & Hawkins, A. R. (1987) Identification and isolation of a putative permease gene in the quinic acid utilisation (*qut*) gene cluster of *Aspergillus nidulans*. *Current Genet.* **12**, 135-139.

Witte, M. M. & Dickson, R. C. (1990) The C_6 zinc finger and adjacent amino acids determine DNA binding specificity and affinity in the yeast activator protein *lac9* and protein *PPR1*. *Mol. Cell. Biol.* **10**, 5128-5137.

Xue, Z. Y., Shan, X. Y., Lapegre, B. & Melese, T. (1993) The amino terminus of mammalian nucleolin specifically recognises SV40 T antigen type nulear localization sequences. *Eur. J. Cell. Biol.* **62**, 13-21.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Yankulov, K., Blau, J., Purton, T., Roberts, S. & Bently, D. L. (1994) Transcriptional elongation by RNA polymerase II is stimulated by transactivators. *Cell* 77, 749-759.

Zhang, L. & Guarente, L. (1994) The yeast activator HAP1, a GAL4 family member, binds DNA in a directly repeated orientation. *Genes Dev.* **8**, 2110-2119.

Zhou, K., Brisco, P. R. G., Hinnkanen, A. E. & Kohlhaw, G. B. (1987) Structure of the regulatory gene *LEU3* and evidence that *LEU3* itself is under general amino acid control. *Nucl. Acids Res.* **15**, 5261-5273.

Zhou, K., Bai, Y. & Kohlhaw, G. B. (1990) Yeast regulatory protein *LEU3* and evidence that *LEU3*: a structure-function analysis. *Nucl. Acids Res.* **18**, 291-298.

Zhou, K. & Kohlhaw, G. B. (1990) Transcriptional activator leu3 of yeast: mapping of transcriptional activation function and significance of activation domain tryptophans. *J. Biol. Chem.* **265**, 17409-17412.