Characterisation of cDNA Clones for the *am* Gene of *Neurospora crassa*

by

CAROL E. LINDSAY

Submitted to the University of Edinburgh as a Thesis required for the degree of Doctor of Philosophy in the Faculty of Science.

Department of Genetics

University of Edinburgh

1991

Abstract

Abstract

ICR 170, an acridine half-mustard, has been shown to induce predominantly frameshift addition mutations in the *am* gene of *Neurospora crassa* (Burns *et al.*, 1986) and other eukaryotic genes. The related compound ICR 191 produces both addition and deletion mutations in prokaryotic DNA. The introduction of the *am* gene into a prokaryote such as *E.coli*, followed by the induction of frameshift mutations using ICR 170 should provide an insight into the mechanism by which frameshift mutations are induced in eukaryotic and prokaryotic DNA and their different responses to induction of frameshift mutations by acridine half-mustards.

In order to study the mechanism by which mutations are induced in prokaryotic and eukaryotic DNA, it is necessary to have a system which will allow the identification of frameshift mutations in a copy of the *N.crassa am* gene which has been transformed into *E.coli*. This study was an attempt to set up such a system. This work involved the production of complementary DNA to the *am* gene with the intention of expressing the *am* gene product, NADP-dependent glutamate dehydrogenase, in an *E.coli* strain which is auxotrophic for glutamate. Expression of the *am* gene product should complement the *E.coli* mutation and revert the *E.coli* strain to glutamate prototrophy. In this study, optimum conditions for producing cDNA from *Neurospora crassa* mRNA were determined and cDNA libraries were produced both from total *Neurospora crassa* mRNA and from *N.crassa* mRNA which had been selected by its ability to hybridise to the *am* genomic sequence.

Potential cDNA clones to the *am* gene were isolated and characterised. 19 clones from the cDNA library constructed using total *N.crassa* mRNA showed homology to the *am* sequence in colony blots. Characterisation of these clones identified them as *am* cDNA clones which contained only the 3' end of the *am* coding sequence. Since all 19 clones represented the same area of the *am* coding sequence, it is likely that some form of structural constraint in the mRNA prevented full length cDNA molecules for *am* being formed.

1328 cDNA clones were identified as homologous to the *am* genomic sequence

within the cDNA library constructed from *am*-selected *N.crassa* mRNA . This represented a 90-fold increase in the representation of cDNA clones to the *am* gene using the selected mRNA.

Characterisation of pAM-7, one of these 1328 cDNA clones by restriction mapping identified the presence of an unexpected *Bam* HI site at the 5' end of the cDNA insert. Further characterisation revealed that pAM-7 contained 5' non-coding sequence and a 5' restriction pattern common to the *am* genomic clone. A number of other cDNA clones which were identified as homologous to the *am* gene showed a similar genomic *am* DNA restriction pattern, suggesting that the cDNA library was contaminated with *am* genomic DNA. Since p8-3, a plasmid containing *am* genomic DNA had been used to select mRNA for use in the construction of this library, the most likely source of *am* genomic DNA contamination was co-elution of p8-3 with the mRNA after hybridisation selection. This was confirmed by the isolation of pAM-8, which was an exact replica of the *am* genomic clone.

Screening of the cDNA library with an oligonucleotide probe specific for the 5' intron in the *am* gene showed that all of the 1328 *am* clones identified in this library represented genomic DNA. This discovery prevented the work being taken any further.

This thesis also considers some of the new developments which have taken place since this work was done, such as the advent of polymerase chain reaction, which can facilitate the construction of cDNA clones for use in expression studies such as this one.

Acknowledgements

I would like to thank all my friends and colleagues in Edinburgh for their encouragement throughout this project.

In particular, I would like to thank Dr. Jane Kinnaird for her invaluable advice during the early part of this work and for making available to me her unpublished results.

This thesis could not have been written without the encouragement and advice offered by Dr. Jeff Bond, so I am especially grateful to him.

I would also like to thank Dr. Henrik Kacser, Sheila Carmichael and Herbert Sauro for providing me with office space and access to a computer terminal while I was writing this thesis and for their help in the production of the tables presented here.

I am completely indebted to Frank Johnston and Graham Brown for producing the photographic figures in this thesis in record time.

I would also like to thank my supervisor, Dr. Brian Kilbey, for allowing me to carry out this work.

Abbreviations

Abbreviations

A_{260nm}	absorbance at 260 nm
AmAc	ammonium acetate
ATP	adenosine triphosphate
BSA	bovine serum albumin
bp	base pairs
cDNA	complementary DNA
CIP	calf intestinal alkaline phosphatase
cpm	counts per minute
CsCl-EtBr	Caesium chloride-Ethidium bromide
CsCl	Caesium chloride
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
(α-^{32}P) dCTP	(α-^{32}P)-2-deoxycytidine triphosphate
DEAE	diethyl amino ethyl
DEP	diethyl pyrocarbonate
dGTP	deoxyguanidine triphosphate
DMSO	dimethyl sulphoxide
dNTP	deoxyribonucleotide triphosphate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dTTP	deoxythymidine triphosphate
DTT	dithiothreitol
E.coli	*Escherichia coli*
EDTA	ethylenediamine-tetra-acetic acid
EtBr	ethidium bromide

GDH	glutamate dehydrogenase
HEPES	[4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid]
HPRI	Human placental ribonuclease inhibitor
IPTG	Isopropyl-β-D-thiogalactopyranoside
KAc	potassium acetate
kB	kilobase(pairs)
MES	4-morpholino-ethane sulphonic acid
MOPS	4-morpholine propane sulphonic acid
mRNA	messenger RNA
NADP(H)	nicotinamide adenine dinucleotide phosphate (reduced)
NADP-GDH	NADP-dependent glutamate dehydrogenase
NaIAc	sodium iodo-acetate
NaAc	sodium acetate
N.crassa	*Neurospora crassa*
$O.D._{260}$	optical density at 260 nm
oligo dT	oligo deoxythymidine
P/C/I	phenol:chloroform:isoamyl alcohol at a volume ratio of 49:49:2
PEG	polyethylene glycol
PIPES	piperazine N-N'-bis(2-ethane sulphonic acid)
poly (A^+)	polyadenylated
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
S.cerevisiae	*Saccharomyces cerevisiae*

S.typhimurium	*Salmonella typhimurium*
SDS	sodium dodecyl sulphate
ssDNA	salmon sperm DNA
tRNA	transfer RNA
Tris-Cl	Tris(hydroxymethyl)aminoethane
UV	ultra violet light
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YE	yeast extract

Table of Contents

Chapter 1 : Introduction

Chapter 1

Introduction

1.1 The Nature of Frameshift Mutations and the Postulated Mechanisms Responsible for Them

Since the genetic code is read in non-overlapping triplets, any insertion or deletion of nucleotides will disrupt the reading frame of the nucleotide sequence unless the number of nucleotides inserted or deleted is a multiple of three. Addition or deletion mutations which result in a disruption of the reading frame are therefore referred to as "frameshift" mutations.

1.1.1 Misaligned Pairing of DNA Results in Frameshift Mutation

In common with other types of mutations, frameshifts have been shown to occur preferentially in certain regions of a nucleic acid sequence, resulting in so-called 'hotspots', and not to occur at all in other areas, the so-called 'silent' areas. Streisinger and coworkers, 1966, showed that spontaneous frameshift mutations occur with high frequency at regions of reiterated bases, an observation which has since been reinforced by the work of Okada *et al.*, 1972; Farabaugh *et al*, 1978; Pribnow *et al.*, 1981; and Streisinger and Owen, 1985.

Based on work done on the lysosyme gene of bacteriophage T4 (Streisinger *et al.*, 1966; Okada *et al.*, 1972; Streisinger and Owen, 1985), Streisinger and coworkers, 1966, 1985, proposed a model to explain why "hotspots" for frameshift mutation should occur at sites of reiterated bases. This model implicates misaligned pairing of the DNA strands during replication as the key mode by which frameshift mutations arise. The model involves the following steps (as shown schematically in figure 1.1) :

1. the creation of a single stranded break, or nick, within one of the DNA strands;
2. digestion of the nicked DNA, stopping adjacent to a strand of reiterated bases;
3. melting of the DNA duplex, which then reanneals with one or more bases out of register;
4. elongation of the DNA using the misaligned strand as primer and the complementary strand as template;
5. replication of the misaligned DNA, leading to one normal, wild-type DNA

molecule and one DNA molecule carrying a frameshift mutation.

Using this 'misaligned pairing' model as a mechanism by which frameshift mutations might arise, there are three factors which determine the frequency and stability of frameshift mutations. These are :

1. the relative stability of the misaligned base in the intermediate structures;
2. the likelihood of DNA synthesis using a primer containing a misaligned base;
3. the probability of excision of the misaligned base.

Structures with a small number of misaligned bases have been shown to be sterically feasible (Fresco and Alberts, 1960), so factor 1 may not be a particular problem. However, taking into consideration factors 2 and 3, stable frameshift mutations should not be a common occurence, which raises an important question : why are some areas 'hotspots' for frameshift mutations, which by definition ought to be a fairly rare event, whereas others would appear to be relatively inert to frameshift mutagenesis?

One explanation for frameshift hotspots may lie in the length of the reiterated sequence. In several examples of frameshift hotspots which involve reiterated sequences, frameshifts appear more likely to occur in regions containing 4 or more reiterated bases (Farabaugh *et al.*, 1978; Pribnow *et al.*, 1981; Streisinger and Owen, 1985). For example, 40 % of spontaneous frameshift mutations in the lysosyme gene of bacteriophage T4 map to four sites which each have a reiteration length of 5 bases. Since the 20 bases responsible for these 4 sites account for only 4 % of the T4 lysosyme gene sequence, these sites can be considered as hotspots for spontaneous frameshift mutagenesis.

So, why are long regions of reiterated bases likely to be hotspots for frameshift mutagenesis?

If the misaligned pairing model is accepted, then obviously the greater the reiteration length, the greater the chance of a melted duplex reannealing incorrectly, since any of the bases in the area of reiteration, with the possible exception of the most 3' base in the sequence, may reanneal out of register. This would account for the results observed with the T4 lysosyme gene.

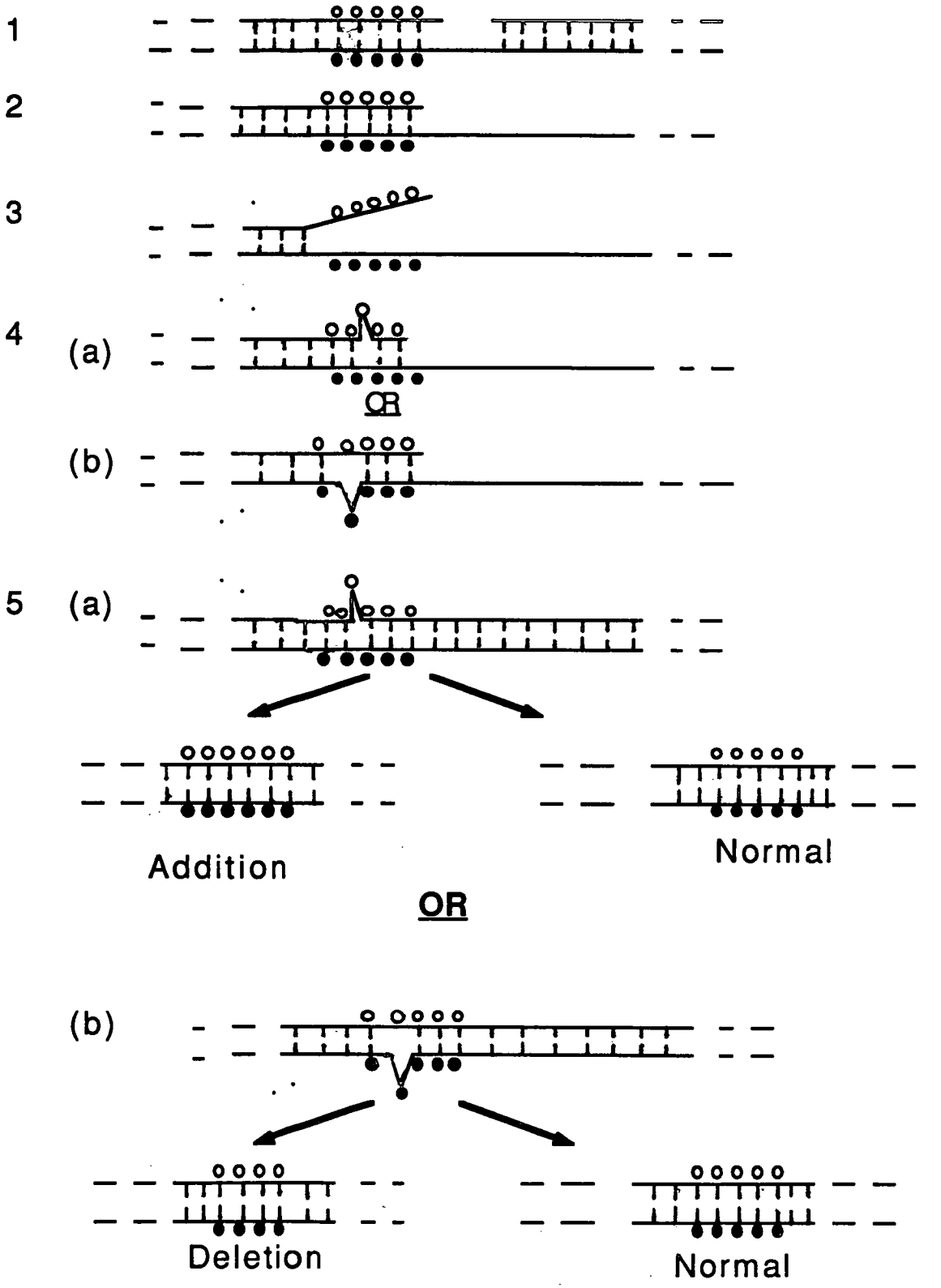
FIGURE 1.1

Misaligned base pairing model for the formation of single base frameshift mutations, postulated by Streisinger and Owen, 1985 (see text for details).

Open circles represent a series of reiterated bases and filled circles represent the complementary bases on the opposite strand.

Note that mispairing of a base on the primer strand (4a and 5a) leads to the formation of an addition mutation; whereas mispairing of a base on the template strand leads to the formation of a deletion mutation.

FIGURE 1.1



A further explanation for the increase in frameshift mutation frequency at regions of long reiteration length may be that long reiteration length could increase the stability of the misaligned base and therefore increase the likelihood of replication of the misaligned stretch of DNA. For example, a misaligned base at the 5' end of a stretch of five reiterated bases would be stabilised by hybridisation between the 4 remaining bases and the complementary bases on the opposite DNA strand. In comparison, a misalignment of the most 5' base in a region of 3 reiterated bases leaves only two bases to hybridise with the complementary bases on the opposing strand and is unlikely to be as stable.

Whatever the explanation, the fact that frameshift mutation hotspots are more likely to occur at long stretches of reiterated bases does support the misaligned pairing model as an explanation of frameshift mutagenesis.

Since it should be equally likely that a base could become misaligned in either the template or the primer strand, and since misalignments in the template strand would lead to frameshift deletions, whereas misalignments in the primer strand would lead to frameshift additions (see figure 1.2), it should be safe to assume that the frequency of frameshift deletion mutations and addition mutations would be equal. However, using the lysosyme gene of bacteriophage T4, Streisinger and Owen, 1985, have observed that the frequency of spontaneous deletion frameshift mutations is typically two to four times greater than that of addition frameshift mutations. So, why should deletion mutations, and therefore misalignment in the template strand, appear to be more stable than addition mutations, and therefore misalignment in the primer strand?

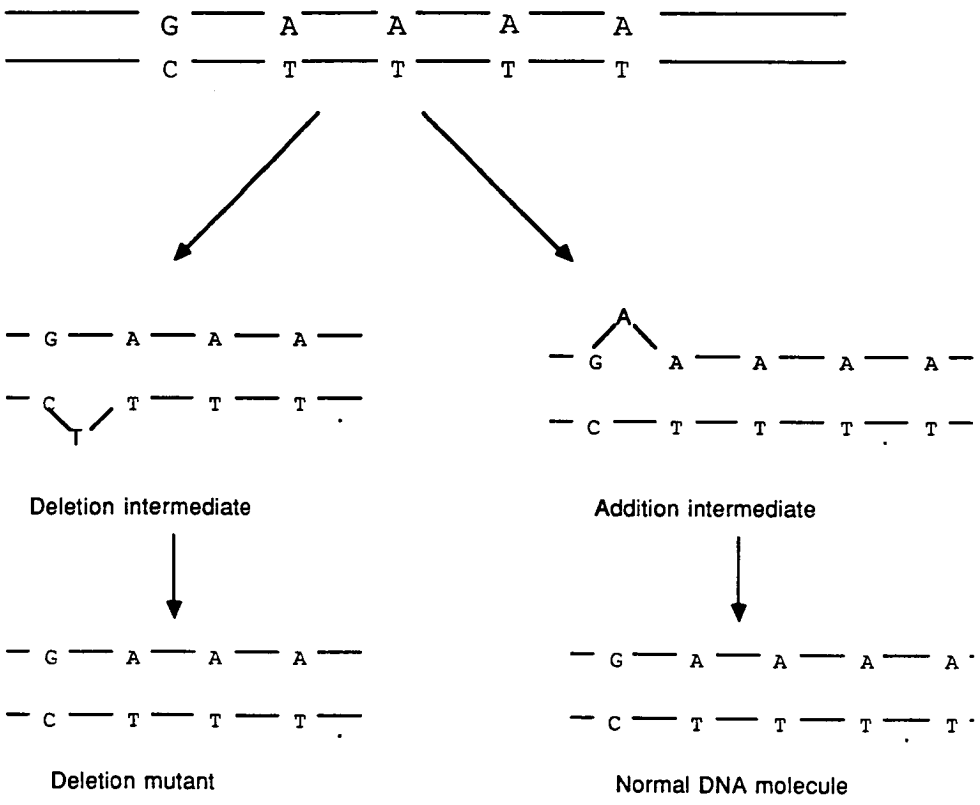
It may be that an extrahelical base in the template strand offers less steric hinderance to the DNA-dependent DNA polymerase during elongation than an extrahelical base on the primer strand; or that an extrahelical base on the primer strand may be more likely to be removed by the 3' exonuclease activity of the polymerase. However, eukaryotic DNA-dependent DNA polymerases do not have a 3' exonuclease activity so the latter explanation cannot apply to them.

A more likely explanation may lie with the DNA repair processes. Consider a repair process which operates by the removal of extrahelical bases. In this case,

FIGURE 1.2

The effect of removal of an extrahelical base by repair enzymes in the formation of a frameshift mutation. Note that removal of an extrahelical base from a deletion intermediate leads to a single base deletion and removal of an extrahelical base from an addition intermediate leads to a normal DNA molecule.

FIGURE 1.2


G A A A A
C T T T T
Deletion intermediate
Addition intermediate
Deletion mutant
Normal DNA molecule

deletions would be favoured since removal of the extrahelical base in a deletion intermediate would stabilise the deletion, whereas removal of an extrahelical base in an addition intermediate would revert the DNA to wild-type, as shown in figure 1.2. Therefore, the more efficient the repair process, the greater the likelihood that a stable frameshift mutation will be a deletion, since any potential additions will be removed prior to replication. If this is the case, then the observed ratio of frameshift deletions to additions seen in the T4 lysosyme gene would appear to support the theory of misaligned pairing as a model to explain frameshift mutagenesis.

As yet, I have only dealt with spontaneous frameshift mutations and frameshift mutation hotspots occurring at regions of reiterated bases. I will deal with frameshift mutations induced by mutagens in a later section of this introduction but I think it is interesting at this point to look at frameshift hotspots which cannot be explained by the misaligned pairing model.

Although many hotspots for frameshift mutation can be accounted for by the misaligned pairing model, there are also frameshift hotspots which do not coincide with regions of reiterated bases. Glickman and Ripley, 1984, have postulated an alternative explanation for spontaneous frameshift mutagenesis which involves DNA secondary structure and this will be dealt with in the next section.

1.1.2 The Role of DNA Secondary Structure in Frameshift Mutation

Among spontaneous deletion mutations of known sequence in the *lacI* gene of *E.coli*, 5 out of 12 have deletion termini in non-reiterated DNA sequences. Analysis of these and neighbouring sequences has revealed the presence of quasipalindromic sequences adjacent to the deletion termini. These quasipalindromic sequences have been implicated in the formation of the deletions. The self-complementarity of palindromic and quasipalindromic sequences may allow the formation of hairpin or cruciform structures which could feasibly bring together the end points of the observed deletions in the *lacI* gene. Excision of the hairpin or cruciform would lead to the specific end points observed (Glickman and Ripley, 1984). This explanation may also account for large deletions, such as the S86 deletion of *lacI*, a 27 base pair deletion which

cannot be explained by reiterated base sequences.

It is also interesting to note that, of the seven spontaneous deletions in the *lacI* gene associated with reiterated sequences, five were also found to be associated with quasipalindromic sequences. This would appear to suggest that, although there are deletion mutations in the *lacI* gene for which only one of the explanations, i.e. either repeats or palindromes, applies, it is likely that the cooperation of the reiterated sequence and the palindromic sequence is preferable for the formation of small deletions. This cooperation may in some way stabilise intermediate structures involved in the formation of the deletion. Perhaps interstrand hydrogen bonding, permitted by the presence of repeats helps to stabilise the hairpins formed by the palindromic or quasipalindromic sequences; or, alternatively, intrastrand hydrogen bonding, permitted by palindromic or quasipalindromic sequences, may stabilise the interstrand misalignments allowed by the repeats.

Glickman and Ripley, 1984, found that 75 % of *lacI* deletions which were explicable by repeats were also explicable by quasipalindromes or palindromes. Interestingly, in the *lacI* gene, only 6.4 % of repeats of 5 or more bases, which allow deletions, coincide with palindromes of 4 bases or more. Similar coincidence has been found for the lysosyme gene of bacteriophage T4 (Owen *et al.*, 1983).

Duplications of DNA have also been associated with palindromic sequences and these may account for the formation of addition mutations at some non-reiterated sites.

From this, it is clear that at least two different mechanisms exist for the production of frameshift mutations and that these mechanisms are not mutually exclusive.

The last two sections have dealt with spontaneous frameshift mutations. In the next section, I will deal with frameshift mutations induced by chemical mutagens.

1.1.3 Induced Frameshift Mutations

Postulated Mechanisms of Action of Mutagens.

It is interesting to note that frameshift mutations induced in response to chemicals show a similar pattern of hotspots to those which occur spontaneously, with frameshifts more likely to occur at regions of reiterated sequence and palindromic sequences. However the spectrum of mutations induced is different from the spontaneously occurring mutations. For example, mutagenesis of the lysosyme gene of bacteriophage T4 using acridine compounds has been shown to increase the frequency of addition mutations by a factor of 10 in reiterated sequences, whereas the frequency of deletion mutations remains unchanged (Streisinger and Owen, 1985). Compare this to the higher frequency of deletions to additions (by a factor of two to four) which occur spontaneously in the same gene, see section 1.1.1. If the mutagen were simply intercalating into the DNA duplex, then the frequencies of addition and deletion mutations as a result of mutagen treatment would be expected to be similar, since the possibility of a base becoming misaligned by intercalation of the mutagen molecule would be the same for each DNA strand. Since this is clearly not the case, simple intercalation cannot be the mechanism by which frameshift mutations are induced in the DNA sequence in response to chemical mutagens. There have been two different explanations put forward to explain the mode of action of a chemical frameshift mutagen on the DNA sequence. These are :

1. the mutagen could stabilise a previously misaligned DNA strand by intercalation

   - within the misaligned stretch of DNA, thereby stabilising the misalignment (Lerman, 1963; Streisinger *et al.*, 1966);
   - opposite the extrahelical base (Lee and Tinoco, 1978);
   - or by stacking of the mutagen molecules on either side of the extrahelical base and thus stabilising it (Streisinger and Owen, 1985).

2. the mutagen could cause the misalignment by intercalation with the primer strand ahead of the replication complex. This could cause the replication complex to stutter and copy the same base more than once (Sakore *et al*, 1979). Alternatively, it may associate with the replication complex in some way, causing it to lose precision (Roth, 1974)

This is shown schematically in figure 1.3.

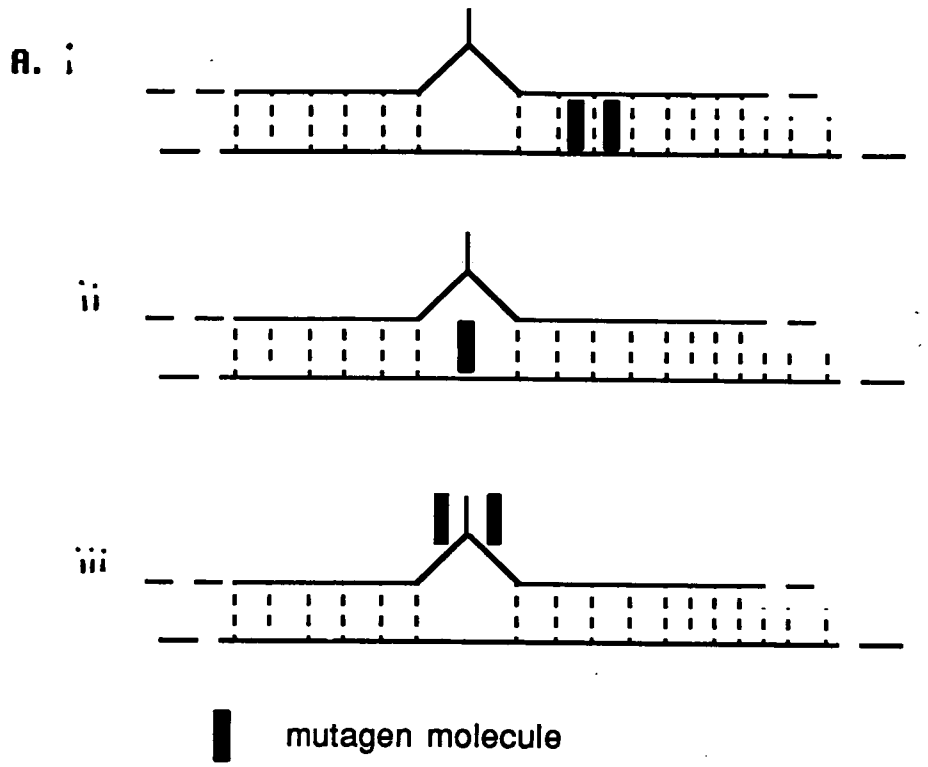
No single model can adequately explain frameshift mutagenesis induced by

FIGURE 1.3

Possible ways in which a mutagen could associate with a DNA molecule in order to stabilise a previous base misalignment within the DNA.

- Ai : intercalation of the mutagen within the duplex
- Aii : Intercalation of the mutagen opposite the extra-helical base
- Aiii: stacking of the mutagens on either side of the misaligned base
- B: Association of the mutagen molecule with a single stranded region of the template strand ahead of the replication complex. this might cause 'stutterring' of the replication complex with the result that a base is copied more than once.

FIGURE 1.3



B

mutagens, since models which explain the induction of frameshifts at reiterated sites do not explain mutations induced at non-reiterated sites and *vice versa*. It is likely that the mutations are generated in a fashion similar to that for spontaneous mutations but that less stable mutations are in some way stabilised or protected from repair by the mutagen. This would explain the large increase in addition mutations produced after chemical induction and the fact that the frequency of frameshift deletions does not change.

Any model invoked to explain the mechanism involved in induced frameshift mutagenesis must account for the following observations :

1. the frequency of mutation is increased in the presence of the mutagen;
2. in addition mutations, the added base(s) reiterates adjacent sequences;
3. the frequency of mutagenesis at any particular area increases as the length of reiteration increases;
4. induced addition frameshift mutations are more frequent than induced deletion frameshift mutations;
5. bulky molecules are capable of inducing mutations;
6. some efficient intercalators are poor mutagens;
7. mutagenesis follows higher-than-first order kinetics.

None of the models proposed for the mechanism of induced frameshift mutagenesis explains all of these criteria, but perhaps the most attractive model is the one which involves the mutagen molecules stacking on either side of the extrahelical base (Streisinger and Owen, 1985). In this model, mutagenesis is due to stabilisation of the extrahelical base by stacking interactions between the base and the mutagen molecules. This model would allow the accommodation of bulky molecules and the higher-than-first order kinetics could be explained by the participation of several mutagen molecules in the "sandwich".

However, this model does not easily account for the higher frequency of additions to deletions, but Streisinger and Owen, 1985, offered the following as possible explanations :

mutagens stacked on an extrahelical base on the template strand, therefore leading to frameshift deletions, may somehow interfere with the operation of the DNA polymerase ;

or, mutagens stacked around an extrahelical base may protect it from repair.

However, it is perhaps more likely that the mutagen is simply serving to stabilise an already occurring mutation, and that the frequency of additions increases because mutations which would normally not be observed, either because of instability or repair, are being stabilised and protected by the presence of the mutagen.

As stated earlier, hotspots exist for chemically induced frameshift mutation, which are similar to those for spontaneous frameshift mutations. However, there are some differences and these will be dealt with in the next section.

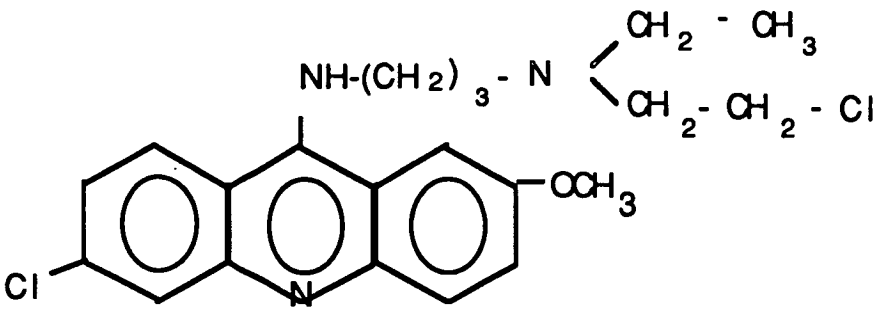
1.1.4 Hotspots for Induced Frameshift Mutagenesis

Using the acridine half-mustard, ICR 191, to induce frameshift mutations in the *lacI* gene of *E.coli*, Calos and Miller, 1981, observed that the induced mutations were predominately additions or deletions of a single G-C base pair at sequences of three reiterated G-C base pairs. Some of the sites mutated within the *lacI* gene were found to predominately produce additions in response to the mutagen, with additions being ten times more common than deletions at these sites; whereas others seemed to predominately produce deletions in response to the mutagen.

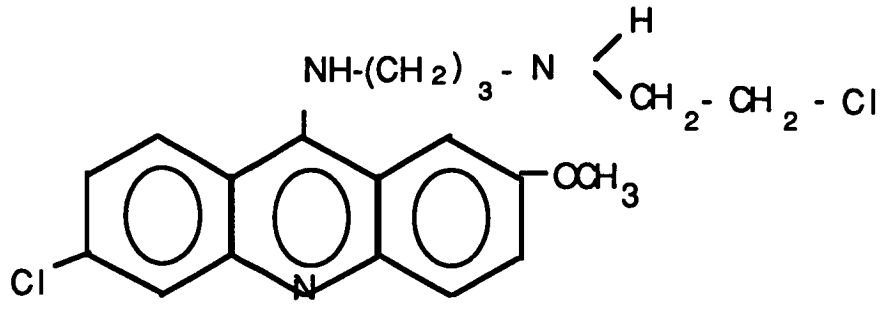
It is possible that ICR 191 interacts preferentially with either G or C, and that the nature of the mutation, i.e. whether a deletion or an addition is produced in response to treatment with the mutagen, at any particular site is determined by which strand contains the favoured base. Therefore, if the favoured base were on the primer strand, additions would be favoured, and similarly, if the favoured base was on the template strand, deletions would be favoured.

Another mechanism for induced frameshift mutagenesis has been proposed by Ernst *et al.*, 1985. This arises from work done on the *CYC 1* gene of the yeast, *Saccharomyces cerevisiae*, using another acridine half-mustard, ICR 170. ICR 170 is similar in structure to ICR 191, see figure 1.4, but for some unknown reason is a more active mutagen in eukaryotic cells, whereas ICR 191 is more active in prokaryotes. ICR 170 induced mutations in yeast are similar in nature to those induced by ICR 191 in *E.coli*, but, whereas ICR 191 induces both additions and deletions of single G-C base pairs at sites of monotonous runs of

FIGURE 1.4



ICR 170



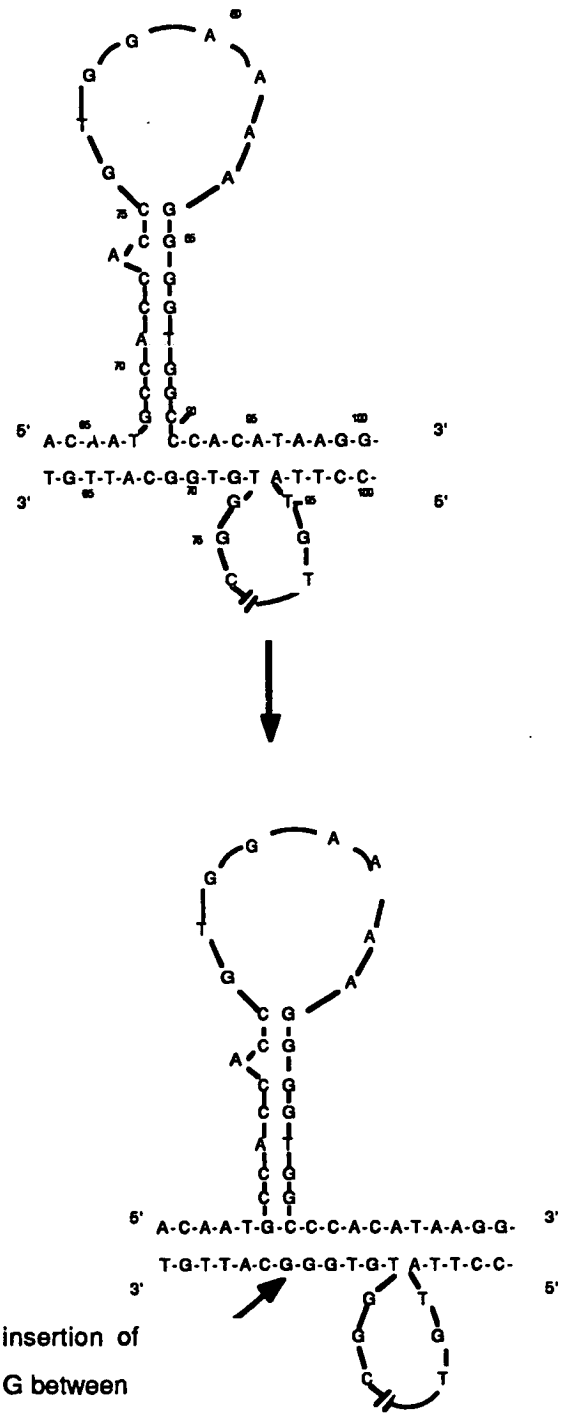
ICR 191

three to four G-Cs in *E.coli*, ICR 170 appears to induce only additions of single G-C base pairs at monotonous runs of two to three G-Cs in the *CYC 1* gene of yeast. The same behaviour is seen when ICR 170 is used to induce frameshift mutations in the *am* gene of *N.crassa*, encoding NADP-dependent glutamate dehydrogenase (Burns *et al.*, 1986).

Ernst *et al*, 1985, noticed that the sequence immediately adjacent to one of the sites in the *CYC 1* gene of yeast, which was highly mutable in response to ICR 170, was capable of permitting the formation of a hairpin type of structure. The site which has the potential to form the hairpin lies between nucleotides 68 and 90, and the mutation hotspot is at nucleotides 90 - 92, where a G-C addition is observed. As shown in figure 1.5, the insertion of a G in the strand of DNA opposite the hairpin would allow further base pairing and further stabilise the hairpin. Following the return of the hairpin to the linear form and subsequent DNA replication or mismatch repair, a G-C addition at position 90 would result. Ernst *et al.* have proposed that ICR 170 promotes this process either by helping in the formation of the hairpin or by stabilising the hairpin and allowing the insertion of G to occur. Although the mechanism proposed by Ernst *et al.*, 1985, cannot account for all of the observed hotspots for frameshift mutation induced by ICR 170 in the *CYC 1* gene of yeast, it may help to partly explain why not all hotspots for frameshift mutation are at reiterated sequences and why only certain monotonous runs of G-C base pairs serve as hotspots for induced frameshift mutagenesis. It also provides a very nice example of cooperation between a site of reiterated sequence with a palindromic sequence as a mechanism for mutagenesis (see section 1.1.2). However, what it does not attempt to explain is why frameshift mutations induced by the acridine half-mustards, ICR 170 and ICR 191, occur preferentially at monotonous runs of G-C base pairs and not at runs of A-T base pairs.

As stated previously, the acridine half-mustards, ICR 170 and ICR 191, although related structurally, appear to have different specificities, with ICR 191 being more active in prokaryotes than eukaryotes and producing both addition and deletion frameshift mutations, whereas ICR 170 is more active in eukaryotes and produces solely addition mutations. Although it should not be ruled out, it

FIGURE 1.5



insertion of G between positions 68 and 69 stabilises the hairpin

unlikely that the difference in structure between two such related mutagen compounds is responsible for their different action in eukaryotes and prokaryotes, or for the different response to the compounds exhibited by the prokaryotes and eukaryotes. It is far more likely that the explanation for the different mutagen specificities observed is a factor of the different cell types involved. For instance, it may be that the different response to mutagens shown by prokaryotic and eukaryotic cells is due to different repair processes invoked in response to the mutagen by the different cell types.

1.2 The *am* Gene of *Neurospora crassa*

In order to study the behaviour of any gene in response to treatment with a chemical mutagen, it is necessary to have a convenient system for identification of the mutants induced. One example of such a system is that for identifying mutations in the *am* gene of *Neurospora crassa* (Kinsey, 1977).

The *am* gene of *Neurospora crassa*, encoding NADP-dependent glutamate dehydrogenase (NADP-GDH), has been the subject of many studies into mutagenesis involving a variety of mutagens (Brett *et al.*, 1976; Kinsey and Fincham, 1979; Kinsey *et al.*, 1980; Siddig *et al.*, 1980; Rambosek and Kinsey, 1984; Burns *et al.*, 1986), and is of particular use in such studies for several reasons :

the gene product, NADP-dependent glutamate dehydrogenase, can be readily identified, both *in vivo*, by its ability to restore glutamate independence to am^- strains, which normally have an absolute dependence on glutamate for growth (Kinnaird *et al.*, 1982; Gurr *et al.*, 1986), and *in vitro*, by assaying for NADP-GDH activity in crude cell extracts, by following the conversion of NADP to NADPH (Coddington *et al.*; 1966, Siddig *et al.*, 1980) in the NADP-GDH catalysed reaction, shown below, by spectrophotometry at 340 nm :

glutamate + $NADP^+$ + $H_2O \rightleftharpoons$ α-ketoglutarate + NADPH + NH_4^+

A comprehensive set of deletion mutants (Rambosek and Kinsey, 1983) is available which allows new mutants to be mapped to particular areas within the

gene.

it is a "housekeeping" gene and therefore the product is likely to be conserved between species. For example, there is 59 % identity at the amino acid level between *N.crassa* NADP-GDH and *E.coli* NADP-GDH, with up to 79 % identity at the region corresponding to amino acids 57 to 175 of the *E.coli* NADP-GDH sequence (McPherson and Wootton, 1983); and 63-64 % homology between *N.crassa* NADP-GDH and *S.cerevisiae* NADP-GDH (Moye *et al.*, 1985, Nagasu and Hall, 1985).

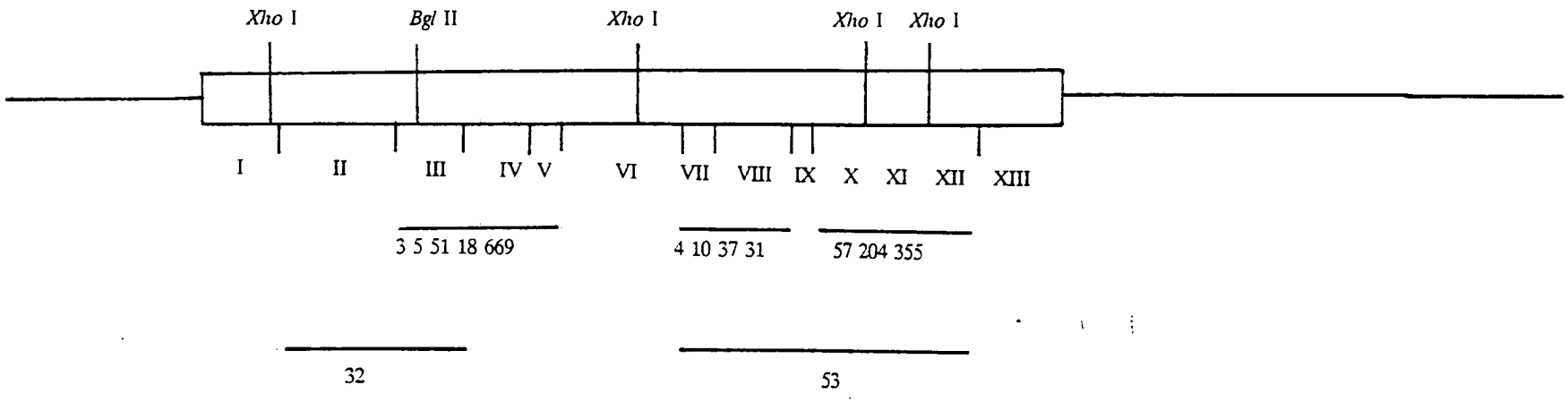
The *am* gene has been isolated from a genomic DNA library (Kinnaird *et al.*, 1982) and sequenced (Kinnaird and Fincham, 1983), and genomic clones are available in both λ bacteriophage and the plasmid vector, pUC 8.

Mutations introduced into the *am* gene can be mapped within the gene by investigating the ability of the mutated DNA to recombine with *am* genes which carry deletions. By crossing a strain of *N.crassa* which carries either a frameshift or a point mutation in *am* with another strain of *N.crassa* which has part of its *am* gene deleted, recombination can occur between the two *am* genes and NADP-GDH activity will be restored, provided the position of the frameshift or point mutation does not lie within the area covered by the deletion. If the position of the frameshift or point mutation does lie within the area covered by the deletion, recombination between the two *am* genes will not lead to a functional *am* gene product. This 'deletion mapping' provides a very useful method of localising new mutations in the *am* gene.

Using this method to identify areas of the *am* gene carrying frameshift mutations following mutagenesis with the acridine half-mustard, ICR 170, Kinnaird and Kilbey produced a deletion map of the induced frameshift mutations (unpublished). Figure 1.6 show the positions of these mutations as calculated from the deletion map (personal communication).

It can be seen quite clearly from this map that not all regions of *am* are equally susceptible to frameshift mutation by ICR 170. For instance, there is an area in the middle of the gene which is apparently 'silent' to mutations, whereas the

FIGURE 1.6



Roman numerals refer to linkage groups designated by Rambosek and Kinsey, 1983.

3, 4, 5, 10, 18, 31, 32, 37, 51, 53, 57, 204, 355 and 669 refer to *Neurospora crassa* mutants identified as carrying frameshift mutations in the *am* gene following treatment with ICR 170 (Jane Kinnaird, personal communication).

areas at either end of the coding sequence appear to be highly mutable with ICR 170. The apparent silent area in the centre of the gene is especially interesting as very few point mutations have been shown to map within that area (Rambosek and Kinsey, 1983).

The mechanisms by which frameshift mutations may have arisen has already been discussed earlier in this introduction. However, it is interesting to note that ICR 170 appears to only produce addition mutations in eukaryotic DNA and not deletion mutations (Donahue *et al.*, 1981; Ernst *et al.*, 1985; Mathison and Culbertson, 1985). This observation is supported by reversion analysis of frameshift mutants of the *am* gene of *Neurospora crassa* (Burns *et al.*, 1986), which shows that *am 15* and *am 129*, both single base frameshift deletion mutants, have a high frequency of reversion when treated with ICR 170. *am 6*, a single base frameshift addition mutation, on the other hand, does not revert when treated with ICR 170, and nor do frameshift mutations originally induced with ICR 170.

Since both additions and deletions occur in response to frameshift mutagenesis by ICR 191, a compound closely related to ICR 170, in *E.coli* and other prokaryotes (Ames *et al.*, 1973; Culbertson *et al.*, 1977; Calos and Miller,1981), it is unlikely that ICR 170 is incapable of producing deletion mutations. It is interesting, therefore, to speculate about what would happen to a eukaryotic gene, such as the *am* gene, if it were subjected to frameshift mutagenesis while it was resident in a prokaryotic environment. For example, if mutation hotspots are determined solely by sequence considerations (Streisinger and Owen, 1985; Glickman and Ripley, 1984), then the effect of induced frameshift mutagenesis should be the same regardless of cellular environment. If, however, the differences in mutability between eukaryotes and prokaryotes are a function of the effectiveness of repair processes in the organisms, (if, for example, eukaryotes are able to repair deletions more effectively than prokaryotes), then both deletions and additions would be expected. In this case, it would be impossible to predict the exact position of the mutations within the gene, except to say that they would be expected to occur at runs of 2 or more GC's, as this is characteristic of frameshift mutations induced by ICR compounds, (Ames *et al.*,

1973; Calos and Miller, 1981; Mathison and Culbertson, 1985; Ernst *et al.*, 1985). Whether they would occur only within the hotspots already identified for ICR 170 induced addition mutations or be more widespread throughout the gene remains to be seen.

A further possible scenario is that proteins normally associated with eukaryotic DNA may shield certain areas of the gene from mutagens leading to the creation of so-called 'silent' areas. If this were the case, then the area in the middle of the *am* gene, apparently silent to mutation, (see fig.1.6) should become mutable, when it is contained within *E.coli* as naked DNA.

Of course, none of these explanations are mutually exclusive and the *in vivo* situation is likely to be a mixture of some or all of these.

1.3 Approaches for Studying Frameshift Mutations Induced in a Eukaryotic Gene Contained in a Prokaryotic Cell

In this study, the system used had to allow, not only a method for identifying the induced mutants, but also a way of identifying the prokaryotic cells which contained the eukaryotic gene prior to mutagenesis. Since mutagenising a eukaryotic gene within a prokaryotic cellular environment was a novel approach, no system was available when this work was undertaken. It was therefore essential that such a system was devised before any frameshift mutagenesis studies could take place.

There are several important criteria which must be met in any system used to study the induction of frameshift mutations in a eukaryotic gene contained within a prokaryotic cellular environment:

1. a suitable prokaryotic host is required, one which will readily accept a foreign piece of DNA and which will be susceptible to mutagenesis.
2. a system for introducing the DNA into the host must be available
3. cells containing mutant copies of the eukaryotic DNA, after the mutagenesis event, must be readily identifiable
4. there must be an easy way of isolating the eukaryotic DNA after mutagenesis and identifying the position of the mutation.

If the *am* gene from *Neurospora crassa* is to be subjected to frameshift mutagenesis while resident within a prokaryotic host cell, there must be a way to identify the presence of the *am* gene within the host cell and to identify cells carrying mutations within *am* after induction. There are two possible ways in which this could be done :

1. directly, by detecting the *am* gene product, glutamate dehydrogenase, by complementation of a glutamate auxotroph of *E.coli*, for example. Frameshift mutations induced in the *am* coding sequence would therefore render the cells auxotrophic for glutamate once again.
2. indirectly, by fusing all or part of the *am* gene sequence to the 5' coding sequence of an *E.coli* gene which encodes a readily identifiable prokaryotic gene product. If the reading frame of the prokaryotic gene and the activity of its gene product can be conserved, mutations induced in the *am* sequence would cause the gene product to lose activity and mutants could be identified.

At this point it is useful to look at each of these screening procedures in more detail. In the first of these procedures, the *am* gene would have to be placed under the control of *E.coli* transcription and translation control sequences and cloned into an *E.coli* strain which is auxotrophic for glutamate. The presence of *am* would be confirmed by the ability of the auxotrophic strain to grow in the absence of externally added glutamate after expression of the *am* gene product. After chemical induction of frameshift mutations in the *am* gene, the cells containing *am* mutants would revert to glutamate auxotrophy and thus could be identified.

The second procedure requires that part or all of the *am* sequence is placed within the reading frame of an *E.coli* gene in such a way as to retain activity of the *E.coli* gene product. Chemical induction of frameshift mutations in *am* would render the *E.coli* gene out of frame and the mutants could be identified by loss of activity of the *E.coli* gene product.

Both procedures are considered in more detail in the following sections.

1.3.1 System One

The first system considered was one which involved studying the entire coding region of the *am* gene by using the gene product, NADP-GDH to complement glutamate auxotrophy in *E.coli* mutants. Since the NADP-GDH from *E.coli* and

the NADP-GDH from *N.crassa* are highly homologous at the amino acid level, expression of *am* in *E.coli* cells which are auxotrophic for glutamate should convert the cells to glutamate independence, by complementation of the mutation in *E.coli*. After mutagenesis, *E.coli* cells containing induced mutations within the *am* gene could be identified by their return to glutamate auxotrophy. This system has the advantage that it considers the gene as a whole and therefore any structural effects, such as hairpins, quasipalindromes, etc. present in the *am* coding sequence, will be taken into account. However, it is dependent on the availability of an appropriate *E.coli* strain, i.e. one which is unable to survive independently without added glutamate, and on the ability of the *am* gene to produce a functional gene product, NADP-GDH, which is capable of complementing glutamate auxotrophy in *E.coli*. The first of these considerations, i.e. the availability of an *E.coli* strain which is auxotrophic for glutamate does not present any problems as there are at least two such strains available: CB100 (Berberich, 1972) which is also referred to as PA340 (Pahel *et al.*, 1978, Sanchez-Pescador *et al.*, 1982); and CLR 207 (Mattaj *et al.*, 1982). Both these strains carry mutations in both the NADP-GDH encoding gene, *gdhA*, and in the gene for glutamate synthase, *gltB*, rendering them glutamate dependent. The *E.coli* strain must be mutated for both glutamate synthase and NADP-GDH in order to make it glutamate dependent as a mutation in only one of the genes does not make the strain auxotrophic for glutamate (Berberich, 1972).

The second consideration, i.e. whether the *am* gene can be expressed in *E.coli* and whether its gene product can complement glutamate auxotrophy in *E.coli* does, however, present a few problems. The *am* gene itself cannot be expressed in *E.coli* due to the presence of two introns, one of 66 base pairs interrupting codon 15 and the other of 61 base pairs interrupting codon 107 (Kinnaird and Fincham, 1983). However, this problem could be overcome by using a cDNA copy of *am* instead of the gene itself. By copying mature mRNA from *N.crassa* into cDNA using reverse transcriptase, the cDNA would not contain introns and could be transcribed and translated correctly in *E.coli* cells, if the correct transcription and translation signals are present on the cDNA. Unfortunately, transcription and translation control sequences for *N.crassa* genes are not sufficiently similar to those of *E.coli* to allow expression of *am* in *E.coli* so it

would be necessary to place the cDNA to the *am* gene under the control of *E.coli* transcription and translation control sequences in order to achieve its expression. This, in itself, does not present any particular problems since there are many systems available for the expression of eukaryotic DNA sequences in *E.coli* using plasmid and bacteriophage vectors.

Once an expression system has been set up for *am* in *E.coli*, there is no guarantee that the *am*-encoded NADP-GDH will be able to perform the same function in *E.coli* and restore glutamate independence to the cells. However, there is a high degree of homology at the amino acid level between the *E.coli* NADP-GDH and the *N.crassa* NADP-GDH, with 59 % amino acid sequence identity, McPherson and Wootton,1983. At this level of homology, there is a very strong chance that the *N.crassa* NADP-GDH will be able to complement the same function in *E.coli* . Moye *et al.*, 1985 were able to isolate the gene which encodes NADP-GDH in *Saccharomyces cerevisiae*, *GDH1*, by its ability to complement the *E.coli* glutamate auxotroph, PA340. NADP-GDH from *S.cerevisiae* shares 59 % homology at the amino acid level with NADP-GDH from *E.coli* and 64 % homology at the amino acid level with the NADP-GDH from *N.crassa* (Nagasu and Hall, 1985), so there is no reason to suggest that the gene product from the *N.crassa am* gene will be incapable of complementing glutamate auxotrophy in *E.coli*.

Assuming that NADP-GDH from *N.crassa* is capable of converting an *E.coli* strain, auxotrophic for glutamate, from auxotrophy to prototrophy, its presence within that *E.coli* strain could be identified by the ability of the auxotrophic strain to grow in the absence of added glutamate. After chemical induction of frameshift mutations in the *am* gene, the cells containing *am* mutants would revert to glutamate auxotrophy and thus could be identified.

However, reversion to glutamate auxotrophy could also be caused if the copy of the *am* cDNA was lost from the cell by a deletion event, so the system must include a way to confirm that the cDNA is still present in the cell after treatment with the mutagen. By introducing the cDNA to the *am* gene into the cells using a vector which carries a selectable marker such as antibiotic resistance and retaining the cDNA on the vector throughout the mutagenesis procedure, the

cells containing the *am* cDNA could be selected for and the *am* cDNA could be easily retrieved from the cells if further analysis was necessary to confirm its presence in the cell. Ease of recovery of the *am* cDNA after mutagenesis would also be desirable to identify the position of the induced mutation after mutagenesis.

1.3.2 System 2

The second system considered for studying frameshift mutations, induced in a eukaryotic gene while it was resident in a prokaryotic host, was to clone all or part of the *am* gene into the 5' coding region of the *E.coli lacZ* gene. The *lacZ* gene of *E.coli* is the system most often used for studies involving in frame fusion of a foreign protein to an *E.coli* protein for several reasons :

- its expression can be carefully controlled, since the *lacZ* gene will only be expressed if grown using lactose as the sole carbon source, or in the presence of a chemical inducer such as IPTG;
- the presence of its gene product, β-galactosidase, within the cell can be identified by the cell's ability to produce dark blue colonies when grown in the presence of a chromogenic substrate, X-gal;
- it can tolerate insertions of foreign DNA into its 5' coding region and, provided that DNA is a multiple of 3 bases in length and does not contain an in frame termination codon, will continue to express active β-galactosidase;
- the gene product, β-galactosidase, is a very large protein and proteins fused to it are often more stable in *E.coli* cells than they would be in their native form. Indeed, for some foreign proteins, it is the only way they can be expressed in *E.coli*.

This system could be used to identify mutations induced in *am* if all or part of *am* could be inserted into the 5' coding region of *lacZ* in such a way as to retain the reading frame of the *lacZ* gene. Since the *lacZ* gene would retain its ability to produce a functional β-galactosidase, clones containing in frame *am* insertions would be blue in colour when grown on the chromogenic substrate, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), after induction with the chemical inducer, isopropyl β-D-thiogalactopyranoside (IPTG).

However there were a number of restraints to using an in frame fusion system in this study. Firstly, the system requires that the *lacZ* gene continues to express a functional β-galactosidase, capable of producing a blue colour in the presence of X-gal, even after the foreign DNA had been introduced into its 5' coding region. Frameshift mutations induced into the *am* DNA would render not only *am* DNA

out of frame, but also the *lacZ* gene and colonies containing a frameshift in *am* would change colour from blue to white. However, there is no way to selectively introduce frameshift mutations into the *am* DNA and mutations introduced into the *lacZ* gene, would also result in a change of phenotype from blue to white, yielding false positive results. In fact, the coding region of the *lacZ* gene is 3074 base pairs long so, even if the entire *am* coding region, which is 1359 bases in length, was included in the *LacZ* sequence, less than one in three of the frameshift mutants identified by a blue to white colour change would contain a mutation in the *am* DNA. One possible solution to this would be to introduce the *am* DNA into a plasmid or bacteriophage vector which carries the *lacZ'* gene containing a polylinker. *lacZ'* encodes the α peptide of β-galactosidase, which is capable of complementing the mutation in *E.coli* mutants which renders them *lac*$^-$ by their inability to produce functional α peptide. By using this, instead of the entire *lacZ* gene for in frame fusion of *am*, then this would increase the chance that a mutant, identified as carrying a frameshift mutation by a blue to white colour change after chemical induction, would contain that mutation within the *am* DNA, rather than the *lacZ* DNA. However, the smaller *lacZ'* gene cannot accept large pieces of DNA and still produce functional β-galactosidase alpha peptide and there is a practical limit of about 100-200 bases which can be incorporated without destroying its α-complementation function. Therefore, small areas of *am* would have to be dealt with in isolation from the rest of the gene sequence, which may affect the way the DNA behaves in response to the mutagen and may not reflect its behaviour *in vivo*.

A further limitation to this system is that it requires DNA to be inserted into the *lacZ'* without disruption of the reading frame and therefore, a way of isolating pieces of DNA from the *am* gene which are multiples of three base pairs in length, and which do not contain an in frame termination codon, must be found. The easiest way to do this would be to introduce restriction fragments of known length from the *am* gene into the polylinker of a plasmid, such as one of the pUC series, or a bacteriophage, such as M13 (Vieira and Messing, 1982; Yanisch-Peron *et al.*, 1985), which contain the polylinker within the *lacZ'* region, at a restriction site which would render the fragment in frame. A further consideration in the choice of fragment to be used is that the piece of *am* DNA

chosen must bear some relevance to the aims of the study in addition to being easily isolated and of appropriate length. A problem which immediately arises is that if, for example, an area of the *am* gene is chosen which covers one of the potential hotspots, identified in figure 1.6, and yields the same mutations as it does when mutated in *Neurospora crassa*, then this may indicate that the position of the mutation is as a result of sequence considerations but, if new mutations arise, then there is no way of determining whether these are as a result of the different cellular environment or because the mutations have been induced into small areas of the gene in isolation from the rest of the gene.

Thirdly, using a system in which recombinant clones have the same phenotype as those carrying non-recombinant vectors, i.e. blue colony or plaque colour, introduces several technical problems. For instance, the only way to differentiate between colonies carrying plasmids with in frame *am* DNA and those carrying plasmids without inserts will be by screening the colonies using radioactive probes made from the *am* gene, or by isolating the plasmid and using restriction endonuclease digestion to determine whether the plasmid is carrying an insert. However, once a colony containing *am* is selected for use in the mutagenesis experiments, there is no way to ensure it retains a copy of the insert. Plasmids which have lost their inserts, but continue to confer a blue phenotype on cells grown in the presence of X-gal, will be difficult to detect without continuous screening using radioactive probes or restriction digestion of isolated plasmid, and, until their detection, may lead to false results.

Since many of these problems would have been very difficult to solve and there was a very strong possibility of generating false results, which would not be detected until the project was at the stage of identifying the induced mutations, using this system, it was decided that this system would not be appropriate for this study. Therefore the approach I decided to take was that described in the previous section, which will be considered in more detail in the next section.

1.4 The Strategy for Using System One

As mentioned earlier, several groups had introduced NADP-GDH genes from *E.coli* (Mattaj *et al.*, 1982; Sanchez-Pescador *et al.*, 1982) and from *S.cerevisiae*

(Moye *et al.*, 1985) into the *E.coli* glutamate auxotroph, PA340 or its derivatives, using a plasmid vector to introduce the DNA into the cells, and had achieved complementation of the glutamate auxotrophy in this manner. It would be essential to use cDNA for the *am* gene because the genomic *am* DNA contains introns. It would also be necessary to use a vector which would supply the appropriate control sequences for transcription and translation of the *am* cDNA in *E.coli* as these were absent from the *am* sequence. Introducing the cDNA for the *am* gene into the 5' end of an *E.coli* gene so that the 5' end of the cDNA sequence is in frame with the *E.coli* gene is the easiest way to provide the *am* cDNA with the appropriate control sequences for transcription and translation. This 'fusion' can then be translated by *E.coli*. This will, however, add a few extra amino acids onto the 5' end of the expressed *am* cDNA but, by keeping the number of extra amino acids to a minimum, this should not interfere with the function of the gene product.

This type of 'fusion' differs from the in frame fusion system already described for system 2 in several ways :

1. it requires only the 5' end of the *am* cDNA to be in frame with the *E.coli* DNA, since the termination codon of the *am* cDNA will act as termination codon for the fusion product, whereas system 2 requires both the 5' and 3' ends of *am* to be fused in frame with the *E.coli* gene and uses the termination codon of the *E.coli* gene as termination codon for the fusion protein;
2. the inserted cDNA does not have to be a multiple of 3 base pairs in length in this system;
3. the reading frame of the *am* cDNA must be conserved using this system or the fusion product will not be able to complement auxotrophy. In system 2, the *am* DNA did not have to retain its reading frame, so long as the length of the DNA was a multiple of 3 base pairs, and there was no termination codon generated when it was inserted into the reading frame of the *E.coli* gene.

As has already been discussed, the *am* gene of *N.crassa* itself cannot be expressed in *E.coli*, due to the presence of two introns, one of 66 base pairs interrupting codon 15 and the other of 61 base pairs interrupting codon 107. However, there is no reason that a cDNA copy of the *am* gene could not be expressed if placed under the appropriate control sequences. The choice of vector and cloning strategy used in this work are outlined in the introduction to Chapter 3.

Chapter 2 : Methods

Chapter 2

Methods

2.1 Restriction Digests

Restriction digests were carried out at 37 °C for 90 minutes in the buffer supplied by the manufacturer, with the exception of *Sma I* digests which were carried out at 25 °C in 1 X *Sma I* buffer. Except where otherwise stated, digests were typically of 1-5 µg of DNA using 10 units of enzyme.

A typical reaction is as follows:

DNA	1 µg
10 X buffer	2 µl
Enzyme	10 units

This was made up to 20 µl with sterile distilled H_2O.

Double digests were carried out sequentially in the same buffer where possible. When this was not possible, an equal volume of 4M NH_4Ac was added after the first digest was complete, followed by an equal volume of isopropanol and the DNA was precipitated at room temperature for 10 minutes. The DNA was recovered by centrifugation in a microcentrifuge for 15 minutes at 4 °C, washed with 70 % ethanol, dried *in vacuo*, then resuspended in 1 X the appropriate restriction buffer for the second enzyme.

2.2 Removal of 5' Terminal Phosphate from pTZ18R after Digestion with the Restriction Endonuclease *Sma* I

50 µg of plasmid pTZ18R DNA was digested with 40 units of restriction endonuclease *Sma* I for 90 minutes at 25 °C. To check that digestion was complete, a sample corresponding to 1 µg of DNA was run on a 1 % agarose minigel containing 0.5 µg/ml ethidium bromide, prepared as described in section 2.13.1. and the DNA visualised under short wave UV light. The presence of a single band at a position corresponding to a molecular weight of 2.8 kB indicated that the reaction had gone to completion.

40 µg of the digested DNA was removed and incubated with 1 unit of calf intestinal alkaline phosphatase for 60 minutes at 50 °C in CIP buffer (50 mM Tris-Cl, pH 9.0; 1 mM $MgCl_2$; 0.1 mM $ZnCl_2$). Following incubation, EGTA was added to a final concentration of 50 mM and the reaction mix was incubated at 65 °C for 45 minutes to stop the phosphatase reaction.

The phosphatased DNA was recovered by phenol extraction, followed by ethanol precipitation at -70 °C for 60 minutes.

The DNA precipitate was recovered by centrifugation for 15 minutes in a microcentrifuge at 4 °C. The pellet was dried in a vacuum dessicator, then redissolved in 20 µl of TE buffer.

2.3 Recovering DNA Fragments from Agarose Gels

DNA fragments, generated by restriction endonuclease digestion were separated by electrophoresis on 1-1.75 % agarose gels, containing 0.5 µg/ml ethidium bromide, and visualised under short wave UV light. A slot was cut in the gel immediately adjacent to the band of interest, and a small square of NA 45 DEAE cellulose paper (Schleicher and Shuell) was inserted into the slot. Electrophoresis was continued until the DNA in the band had become completely bound to the NA 45 paper. The paper was removed from the gel, rinsed several times with TE buffer then placed in a sterile Eppendorf tube. Enough DEAE elution buffer was added to cover the paper and the tube was incubated at 60 °C for at least 60 minutes. The paper was removed from the tube and discarded and the buffer, containing the eluted DNA, was diluted to 0.5 M NaCl and centrifuged for 3 minutes on a microcentrifuge, to spin down any insoluble particles which may have been picked up from the agarose gel. The DNA in elution buffer was then transferred to a fresh sterile Eppendorf tube and precipitated in 2 volumes of ethanol at -70 °C for 40 minutes. The DNA was recovered by centrifugation in a microcentrifuge for 15 minutes at 4 °C.

The DNA pellet was dried *in vacuo*, then resuspended in 20 µl of TE buffer. The concentration of the DNA was established by spotting a 1 µl sample onto a 1 % agarose plate containing 0.5 µg/ml ethidium bromide, as described in section

2.4 Determining the Concentration of Small DNA Samples Using Ethidium Bromide-Agarose plates

20 ml of 1 % agarose in Tris-Borate containing 0.5 µg/ml ethidium bromide was made up as described in section 2.13.1., poured into a disposable petri dish, and allowed to set at room temperature.

1 kB ladder (Gibco BRL) was diluted to the following concentrations : 200 ng/µl; 100 ng/µl; 50 ng/µl; 25 ng/µl; 20 ng/µl; 10 ng/µl; and 5 ng/µl, and 1 µl samples of each of these dilutions was spotted onto the EtBr-agarose plate.

A 1 µl sample of the DNA to be tested was spotted onto the plate and the plate was allowed to stand at room temperature for 30 minutes. The DNA was visualised under short wave UV light and the quantity of the DNA samples estimated by comparison with the standards of known concentration.

2.5 Preparation of Plasmid DNA

2.5.1 Plasmid Miniprep.

This was carried out according to the lysis by boiling method detailed in Maniatis *et al.*, 1982.

A single colony of *E.coli* cells containing the plasmid of interest was used to innoculate 5 ml of LB medium containing 75 µg/ml ampicillin, and incubated at 37 ^{o}C overnight with constant shaking. 1.5 ml of this culture was transferred to an Eppendorf tube and the cells were recovered by centrifugation in a microcentrifuge for 45 seconds. The supernatant was removed and the cells resuspended in 200 µl of STET solution on ice. 20 µl of a 10 mg/ml solution of lysosyme was added and the tube was placed in a vigorously boiling water bath for 1 minute.

The cell debris was pelleted by centrifugation in a microcentrifuge for 10 minutes at room temperature and the pellet was removed using a toothpick. 20 µl of 3 M

NaAc was added, followed by 200 µl of isopropanol and the tube was placed at -20 °C for 10 minutes.

The precipitate was recovered by centrifugation in a microcentrifuge for 5 minutes and the pellet was washed with 70 % ethanol, dried *in vacuo* and resuspended in 50 µl of TE buffer.

An equal volume of P/C/I was added, mixed to form an emulsion, and the mixture was centrifuged in a microcentrifuge for 1 minute to separate the phases. The upper aqueous phase was retained and re-extracted with P/C/I until the interphase was clean. 2 volumes of diethyl ether was added to the final aqueous phase, mixed, and the phases separated by centrifugation in a microcentrifuge for several seconds. The ether extraction was repeated, 1/10 volume of 3 M NaAc was added, followed by 2 volumes of absolute ethanol, and the DNA was precipitated at -20 °C overnight, or at -70 °C for 40 minutes. The precipitate was recovered by centrifugation in a microcentrifuge for 15 minutes at 4 °C, dried *in vacuo*, and resuspended in 15 µl of TE buffer. 1 µl of the miniprep DNA was used in restriction digests.

2.5.2 Plasmid MidiPrep.

50 ml of LB medium, containing 75 µg/µl ampicillin, was innoculated with a single colony of the *E.coli* strain containing the plasmid of interest. The culture was incubated overnight at 37 °C with constant shaking.

The cells were transferred to a 50 ml polypropylene centrifuge tube and harvested by centrifugation in an 8 X 50 ml rotor in a Sorvall Superspeed RC-2B centrifuge for 10 minutes at 7000 rpm, 4 °C. The harvested cells were resuspended in 3.5 ml of 25 mM Tris-Cl, pH 8.0; 10 mM EDTA; 15 % sucrose and placed on ice. 0.5 ml of fresh lysosyme (16 mg/ml in 25 mM Tris-Cl, pH 8.0; 10 mM EDTA; 15 % sucrose) was added and the mixture incubated on ice for 20-40 minutes.

8 ml of fresh 0.2 M NaOH; 1 % SDS was added and the mixture incubated on ice for 10 minutes, prior to the addition of 5 ml of 3 M NaAc, pH 5.2 and a further 10 minute incubation on ice.

The cell debris was pelletted by centrifugation in an 8 X 50 ml rotor in a Sorvall Superspeed RC-2B centrifuge for 15 minutes at 10,000 rpm, 4 °C. 20 µl of prebolled λRNase (10 mg/ml) was added to the supernatant and this was incubated at 37 °C for 20 minutes. An equal volume of P/C/I (at a volume ratio of 49:49:2) was added and mixed by inversion. The phases were separated by centrifugation for 20 minutes at 8000 rpm, 4 °C. The upper aqueous phase was removed, 1/10 volume of 3 M NaAc, pH 5.2 was added and the DNA was precipitated with 2 volumes of ethanol at -70 °C for 40 minutes. The DNA precipitate was recovered by centrifugation for 40 minutes at 12,000 rpm, 4 °C, and resuspended in 0.5 ml H_2O.

The DNA was transferred to an Eppendorf tube, 5 µl of RNase (10 mg/ml) was added and the DNA solution incubated at 37 °C for 20 minutes. The DNA was then phenol extracted 2-3 times until the interphase was clean. After phenol extraction, the DNA was precipitated in ethanol at -20 °C overnight. The DNA precipitate was recovered by centrifugation for 15 minutes in an microcentrifuge at 4 °C, dried briefly *in vacuo*, and resuspended in 200 µl of TE buffer. The absorbance of the DNA at 260 nm was determined and the concentration of the DNA calculated. The DNA concentration was adjusted to approximately 1 mg/ml and the purity of the DNA checked by restriction analysis and agarose gel electrophoresis.

2.5.3 Plasmid Maxiprep.

This was essentially the Alkaline Lysis method described in Maniatis *et al.*, 1982. A colony of cells carrying the plasmid of interest was used to innoculate 25 ml of LB or LM liquid medium containing 75 µg/µl ampicillin. This was incubated at 37 °C for 12 hours with constant shaking. This culture was then used to innoculate 2 X 500 ml of LB (or LM), containing ampicillin at a concentration of 75 µg/µl, and this was incubated at 37 °C overnight with constant shaking.

The cells were harvested by centrifugation at 5000 rpm for 10 minutes at 4 °C in a Sorvall Superspeed RC-2B centrifuge. The supernatant was discarded and the pellet washed in 100 ml of ice-cold STE/500 ml of original culture. The pellet was recovered by centrifugation as before, then resuspended in 10 ml of solution I (per 500 ml of original culture) containing 5 mg/ml lysosyme.

*In an 8 X 50 ml rotor in a Sorvall Superspeed RC-2B centrifuge.

The mixture was transferred to 50 ml polypropylene centrifuge tubes and allowed to stand at room temperature for 5 minutes. 20 ml of fresh solution II (per 500 ml of original culture) was added and the contents of the tube mixed by inversion. The tubes were placed on ice and chilled for 10 minutes. 15 ml of ice-cold 5 M KAc, pH 4.8 (per 500 ml of original culture) was added, mixed and the mixture incubated on ice for 10 minutes. The mixture was then centrifuged at 12000 rpm for 40 minutes at 4 °C, and equal quantities of the supernatant transferred to 30 ml Corex tubes. 0.6 volumes of isopropanol was added to each tube and the mixture incubated at room temperature for 15 minutes. The DNA was recovered by centrifugation at 12000 rpm for 30 minutes at room temperature. The supernatant was discarded and the pellet washed with 70 % ethanol at room temperature. The DNA pellet was dried *in vacuo* and dissolved in 8 ml of TE buffer, pH 8.0, (per 500 ml of original culture).

2.5.4 Purification of Closed Circular DNA by Centrifugation to Equilibrium in Caesium Chloride - Ethidium Bromide Gradients.

This was as described in Maniatis *et al.*, 1982.

1 g of CsCl was added per ml of DNA solution and mixed until all the salt had dissolved. 0.8 ml of a 10 mg/ml ethidium bromide solution was added per 10 ml of CsCl solution, and mixed. The mixture was transferred to an ultracentrifuge tube and centrifuged for 40 hours at 35,000 rpm, 20 °C. The plasmid band was visualised under short wave UV light and removed using a syringe, as described in Maniatis *et al.*, 1982. Ethidium bromide was removed by adding 2 volumes of isoamyl alcohol, mixing by inversion over 5 minutes, and then separating the phases by centrifugation for 3 minutes in a microcentrifuge. The upper amyl alcohol layer was removed and amyl alcohol extractions repeated a further 3-6 times until all the ethidium bromide was removed. The aqueous phase was dialysed against several changes of TE buffer, pH 8.0, to remove CsCl. 1/10 of a volume of 3 M NaAc was added and the DNA was precipitated in 2 volumes of ethanol at -20 °C overnight. The DNA precipitate was recovered by centrifugation for 15 minutes in an Eppendorf centrifuge at 4 °C. The pellet was dried *in vacuo*, then resuspended in TE buffer. An absorption spectrum of the DNA was run from A_{200nm} to A_{350nm} and the concentration of the DNA sample calculated, from its absorbance at A_{260nm}. (If the ratio of

A_{280nm}:A_{260nm} was less than 2, then the DNA was phenol extracted and ethanol precipitated to remove residual protein contamination. The purified plasmid DNA was checked by restriction analysis and agarose gel electrophoresis.

2.5.5 Purifying Plasmid Maxiprep on Sepharose 2B

A 30 X 1 cm column of Sepharose 2B was set up and equilibrated overnight with 0.3 M NaCl; 50 mM Tris-Cl, pH 7.5.

The effluent tube from the column was connected to a peristaltic pump via a Uvichord spectrometer set at 260 nm. A chart recorder was connected to the Uvichord.

Up to 2.5 mg of plasmid, which had been centrifuged through a CsCl-EtBr gradient, in 1.5 ml of TE buffer was applied to the column with the pump switched on. The column was then eluted using 0.3 M NaCl; 50 mM Tris-Cl, pH 7.5. The absorption peak was collected in 1 ml fractions and the $O.D._{260}$ of each fraction was determined. The fractions with the highest $O.D._{260}$ readings were pooled and ethanol precipitated. The DNA precipitate was recovered by centrifugation for 20 minutes at 8000 rpm, 4 ^{o}C, dried *in vacuo*, and resuspended in an appropriate volume of TE buffer. The DNA concentration was calculated by measuring the absorbance of the plasmid preparation at 260 nm.

2.6 Preparation and Transformation of Competent Cells

2.6.1 Preparation of Competent Cells

This method is a modification of that described by Hanahan, 1983.

E.coli strain NM522, Gough and Murray, 1983, was maintained as colonies on minimal M9 agar plates. A single colony of NM522 was used to innoculate 10 ml of SOB medium. This culture was incubated overnight at 37 ^{o}C without shaking.

0.5 ml of this overnight culture was used to innoculate a flask containing 30 ml of SOB medium, and this culture was incubated at 37 ^{o}C with constant shaking until the $O.D._{600}$ of the culture reached 0.3-0.4. The culture was then transferred to a sterile 50 ml polypropylene centrifuge tube and left on ice for 30-60 minutes.

The cells were harvested by centrifugation using an 8 X 50 ml rotor in a Sorvall Superspeed RC-2B centrifuge at 7000 rpm for 10 minutes at 4 ^{o}C. The supernatant was discarded and the pellet resuspended in 10 ml of TFB. The resuspended cells were left on ice for 15-30 minutes, then harvested by centrifugation as before.

The pellet from this spin was resuspended in 2 ml of TFB. 70 µl of fresh DMSO was added and the mixture incubated on ice for 5 minutes. 155 µl of 1 M DTT was then added and the mixture incubated on ice for 10 minutes. Finally, a further 70 µl of DMSO was added and the mixture was left on ice for a further 5 minutes.

210 µl samples of the competent cells were then placed in chilled Eppendorf tubes, ready for transformation.

2.6.2 Transformation of Competent Cells

10 ng of uncut pTZ18R plasmid DNA or half of the cDNA ligation mix was added to the cells and the cells were incubated on ice for 30 minutes. The cells were heat shocked at 42 ^{o}C for 60 seconds, then placed on ice for 1-2 minutes. SOC medium was added to give a final volume of 1 ml and the cells were incubated at 37 ^{o}C for 50-60 minutes, with gentle shaking. The cells were then serially diluted using SOC and plated on LM agar plates containing ampicillin at a concentration of 75 µg/ml. Where required, 7 µl/ml of a 20 mg/ml stock IPTG solution and 10 µl/ml of a 20 mg/ml stock solution of X-gal was added to the plates. The plates were incubated overnight until distinct colonies of about 1-2 mm in size were evident. Colonies were then counted and transformation efficiencies calculated.

2.6.3 Calcium Chloride Transformation of Competent Cells

This method was as described in Maniatis *et al.*, 1982. The *E.coli* cells were made competent as follows. A single *E.coli* colony was used to innoculate 10 ml of LB medium and the culture was incubated overnight at 37 ^{o}C. 0.5 ml of the overnight culture was used to innoculate 30 ml of LB medium and this was incubated at 37 ^{o}C until the $O.D_{600nm}$ was 0.3-0.4. The cells were recovered by centrifugation at 7000 rpm for 10 minutes at 4 ^{o}C in a Sorvall Superspeed RC-2B

centrifuge. The cells were resuspended in 10 ml of ice-cold 50 mM $CaCl_2$ and incubated on ice for 30 minutes. After incubation, the cells were recovered by centrifugation at 7000 rpm for 10 minutes at 4 °C, then resuspended in 2 ml of 50 mM $CaCl_2$. 300 µl aliquots of the competent cells were placed in ice-cold Eppendorf tubes and 10 ng of plasmid DNA was added to the cells. The cells were incubated on ice for 30 minutes and then placed at 42 °C for 2 minutes. Following this heat shock step, 1 ml of LB medium was added and the tubes were incubated at 37 °C for 45 minutes without shaking. 100 µl samples of the transformed cells were spread on LB plates containing ampicillin at a concentration of 75 µg/ml. Where required, 7 µl/ml of a 20 mg/ml stock IPTG solution and 10 µl/ml of a 20 mg/ml stock solution of X-gal was added to the plates. The plates were incubated overnight until distinct colonies of about 1-2 mm in size were evident. Colonies were then counted and transformation efficiencies calculated.

2.7 Storage of *E.coli* Colonies on Nitrocellulose

E.coli colonies were stored on nitrocellulose by a modification of the method described by Hanahan and Meselson, 1980.

Colonies were grown on LB agar at 37 °C until they reached a diameter of approximately 2 - 3 mm. A sterile nitrocellulose disc was placed on top of the agar plate and left for 30 seconds. This disc was then transferred, colony side uppermost, to a fresh F plate and incubated at 37 °C for 2 hours.

After incubation, a second sterile nitrocellulose disc was placed on top of the first and the discs were pressed together using a sterile velvet replicating tool, and keyed to each other by punching a series of holes through them using a sterile needle. This "sandwich" was then either placed between damp sheets of sterile Whatmann 3MM paper, sealed in plastic and stored at -70 °C until use, or separated and the replica disc placed colony side up on LB agar and incubated at 37 °C to allow the colonies to develop. The colonies on the replica could then be lysed and their liberated DNA probed with a radioactive probe, as described in sections 2.16.2 and 2.16.3.

A second sterile nitrocellulose disc was keyed to the master disc and this sandwich was then stored at -70 °C. Discs which had been stored at -70 °C were allowed to thaw to room temperature before being separated and allowed to grow on LB. The master disc was placed colony side up on a fresh F plate and incubated at 37 °C until the colonies had recovered before another replica was made.

2.8 Isolation of RNA from *Neurospora crassa*.

Strain of *Neurospora* used was ST-A, the phenotype of which is wild type.

Total RNA was isolated from *Neurospora crassa* by the method of Kinnaird and Fincham, 1983.

10 agar slants each containing 5 ml of Vogel's minimal agar were prepared and innoculated with conidia using a sterile loop. These were incubated at 25 °C until conidia had formed.

The conidia were harvested by suspension in 2 X 5 ml of sterile distilled water and transferred to a sterile flask. The conidial suspension was then filtered through sterile muslim and the filtered suspension used to innoculate 6 flasks, each containing 500 ml of Vogel's minimal medium. These flasks were incubated at 30 °C with constant shaking for 24 hours, until mycelia were produced.

The mycelia were harvested by filtration under vacuum and the filtered mycelia freeze-dried.

2-3 g of the lyophilised cells were finely powdered using a mortar and pestle. The rest of the cells were stored, sealed in plastic, at -20 °C for future use. The powdered cells were resuspended with vigorous stirring in 40 ml of a 1:1 mixture of buffer A and phenol solution (water-saturated phenol:chloroform:isoamyl alcohol at a volume ratio of 49:49:2). This mixture was centrifuged* at 8000 rpm for 20 minutes at 4 °C to separate the phases. The upper aqueous phase was retained and re-extracted four times with fresh phenol solution.

The interphase from the first phenol extraction was back-extracted with 5 ml of

*In an 8 X 50 ml rotor in a Sorvall Superspeed RC-2B centrifuge.

buffer A, and phenol-extracted three times. The final aqueous phase from this was combined with the main aqueous phase, and the RNA in the combined aqueous phase precipitated with 2 volumes of ethanol overnight at -20 ^{o}C.

The precipitate was recovered by*centrifugation at 12000 rpm for 5 minutes at 4 ^{o}C.

The supernatant was discarded and the pellet resuspended in 1 mM EDTA to a volume which resulted in a total absorbance at 260 nm of less than 50 O.D. units.

The solution was then brought to 0.5 M NaAc (pH 5.0), 0.5 % SDS, 1 mM EDTA and extracted a further twice with an equal volume of phenol solution.

The final aqueous phase was brought to 0.2 M NaAc and precipitated overnight with 2 volumes of ethanol at -20 ^{o}C.

A 5 ml sample of the ethanol precipitate was *centrifuged at 12000 rpm for 30 minutes at 4 ^{o}C and the pellet resuspended in 2 ml of sterile distilled water. The absorbance of this solution at 260 nm was measured and the concentration of RNA calculated.

The quality of the RNA was checked by running a sample, equivalent to 5-10 µg of RNA, on a 1.2 % agarose gel with formaldehyde at a concentration of 2.2 M, as described in section 2.13.2.

2.9 Isolation of Poly (A^+) mRNA.

A 10 ml Biorad column was used for the isolation of poly (A^+) mRNA.

The column and tubing were sterilised using 25 ml of a 0.1 % solution of DEP, and then rinsed with 100 ml of sterile distilled water to remove any traces of DEP, prior to the addition of oligo dT cellulose.

0.4 g of oligo dT cellulose in Binding Buffer was added to the column and washed with 25 ml of 0.1 M NaOH; 50 ml of sterile distilled water; and 25 ml of Binding Buffer.

*In an 8 X 50 ml rotor in a Sorvall Superspeed RC-2B centrifuge.

A sample containing 8 mg of total RNA (prepared as described in section 2.8) was *centrifuged at 12000 rpm for 30 minutes at 4 °C. The resulting pellet was dried *in vacuo*, and the dried pellet resuspended in Binding Buffer prior to heating to 65 °C for 10 minutes to denature any secondary structure which might be present.

The RNA solution was passed through the column five times, and the oligo dT cellulose then washed extensively with Binding Buffer until the absorbance of the eluant at 260 nm was less than 0.1 O.D. unit.

The bound poly (A^+) RNA was eluted using Eluting Buffer and the eluant collected in 40 drop fractions using a fraction collector. The absorbance of each fraction at 260 nm was determined, and the fractions containing the absorbance peak were pooled.

The pooled fractions were brought to 0.2 M NaAc and the RNA was precipitated with 2.5 volumes of ethanol at -20 °C overnight.

The precipitate was *centrifuged at 12000 rpm for 30 minutes at 4 °C. The supernatant was discarded and the pellet dried briefly in a vacuum dessicator at room temperature, prior to being resuspended in 2 ml of 0.2 M NaAc and precipitated with 2.5 volumes of ethanol at -70 °C for 40 minutes. This step was repeated a further twice.

The final pellet was dried *in vacuo* and taken up in a small volume of sterile distilled water to give an approximate poly (A^+) RNA concentration of 1 mg/ml, as calculated from the O.D. reading of the sample at 260nm.

A 5 µg sample of the poly (A^+) RNA was run on a 1.2 % agarose gel with 2.2 M formaldehyde to check the integrity of the RNA.

2.10 Hybridisation Selection of *am* mRNA from total *Neurospora crassa* RNA.

This method was based on the hybridisation selection procedures described by Ricciardi *et al*, 1979, and Parnes *et al*, 1981.

*In an 8 X 50 ml rotor in a Sorvall Superspeed RC-2B centrifuge.

2.10.1 Preparation of Filters.

40 µg of a 1 µg/µl solution of plasmid p8-3 (a pUC 8 plasmid derivative containing the 2.7 kB *Bam* HI fragment of *Neurospora crassa* genomic DNA, which contains the entire *am* gene (Kinnaird and Fincham, 1983), was boiled for 10 minutes, then cooled rapidly on ice. An equal volume of 1 M NaOH was added and the mixture incubated at room temperature for 20 minutes.

Following incubation, a 0.5 volume of a solution of 10 X SSC; 0.5 M Tris-Cl, pH 8.0; 1 M HCl was added and the resulting mixture chilled on ice.

The DNA solution was then spotted onto 2 circles (1 cm radius) of Genescreen plus in 5 µl samples until 20 µg of DNA was spotted onto each filter.

The filters were air dried for 1 hour, then placed in a sterile glass tube and washed extensively with 6 X SSC. After allowing the filters to dry in air, the dry filters were placed in a second sterile glass tube, which had a loose fitting metal cap, and baked for two hours at 80 °C under vacuum.

2.10.2 Hybridisation of Filters.

The baked filters were placed in a sterile Eppendorf tube and hybridised in 0.5 ml of hybridisation solution which had been heated to 70 °C for 10 minutes. The hybridisation solution was as follows :

65 %	formamide
20 mM	PIPES
0.2 %	SDS
0.4 M	NaCl
75 µg	total *Neurospora* mRNA (prepared as described in sections 2.8 and 2.9)

This was made up to 0.5 ml using sterile distilled water.

The filters were hybridised for two hours at 50 °C in a shaking platform incubator.

Following hybridisation, the hybridisation solution was removed by aspiration and the filters washed 9 times with Wash Buffer A, then twice with Wash Buffer B. Both wash buffers were maintained at 65 °C throughout this process. After the addition of each wash, the tube containing the filters was

vortexed briefly and the wash buffer removed by aspiration.

After washing, the filters were transferred to a fresh sterile Eppendorf tube, 500 µl of sterile distilled water was added, followed by 50 µl of yeast tRNA (1 µg/µl). The filters were boiled for 1 minute then quick-frozen by placing in a saline ice bath and storing at -70 °C for 15 minutes. The solution was then thawed at room temperature and the filters removed. The filters were dried and stored at 4 °C in a sterile Eppendorf tube.

The solution containing the eluted RNA was made to 0.2 M NaAc and the RNA precipitated in 2.5 volumes of ethanol at -20 °C overnight.

The precipitate was recovered by centrifugation in a microcentrifuge for 10 minutes at 4 °C and the resulting pellet was washed twice with 90 % ethanol. The pellet was dried in a vacuum dessicator, resuspended in 4 µl of sterile distilled water, and stored at -70 °C.

2.11 Preparation of cDNA.

2.11.1 cDNA Synthesis.

cDNA synthesis was carried out using the cDNA synthesis system from Amersham International p.l.c., according to the manufacturer's instructions. The system is based on the method for cDNA synthesis described by Gubler and Hoffman,1983 and uses RNase H and *E.coli* polymerase I, for synthesis of the second strand.

The first strand synthesis reaction was set up in a 1.5 ml Eppendorf tube on ice. The reaction mix was as follows :

5 X 1st strand reaction buffer	4 µl
Sodium pyrophosphate solution	1 µl
Human Placental Ribonuclease Inhibitor (HPRI)	1 µl
dNTP mix	2 µl
oligo dT primer (14-17 mer)	1 µl
(α-^{32}P)-dCTP	5 µCi
Poly (A^+) mRNA(1 µg)	1 µl

This was made up to 19 µl with RNase-free sterile distilled water.

The reaction mix was vortexed briefly, after which 1 µl of reverse transcriptase (20 units) was added , and the first strand synthesis reaction mix was incubated at 42 °C for 60 minutes.

The reaction mix was then placed on ice and 2 X 1 µl samples were removed. These samples were used to estimate the incorporation of radioactivity into nucleic acid. A sample representing 10000-30000 cpm incorporated into nucleic acid was also removed and used as a sample for an alkaline agarose gel, as described in sections 2.13.3 and 2.13.4.

The second strand synthesis reaction components were added to the first strand reaction mix as follows:

1st strand reaction mix	20.0 µl
2nd strand reaction buffer	37.5 µl
(α-^{32}P)-dCTP	50.0 µCi
E.coli RNase H (0.8 units)	1.0 µl
E.coli DNA Polymerase I (23 units)	6.5 µl

This was made up to 99.5 µl with RNase-free sterile distilled water.

After mixing, the second strand reaction mix was incubated sequentially as follows:

1. 60 minutes at 12 °C
2. 60 minutes at 22 °C
3. 10 minutes at 70 °C

After the final incubation, the reaction mix was placed on ice, 0.5 µl of T4 DNA Polymerase (2 units) was added and the total reaction mix incubated for 10 minutes at 37 °C.

The reaction was stopped by the addition of 10 µl of 0.25 M EDTA, pH 8.0, and 10 µl of 10 % SDS.

2 X 1 µl samples were removed and used to estimate the incorporation of radioactivity and the percentage of second strand cDNA synthesised from first strand, as described in section 2.11.3.

A third sample representing 10000-30000 cpm incorporated into nucleic acid was

also removed and used as a sample for an alkaline agarose gel, as described in section 2.11.3.

The double stranded cDNA was purified by phenol/chloroform extraction and ethanol precipitation, as described in section 2.11.2.

2.11.2 Purification of Double Stranded cDNA

(a). Phenol Extraction.

An equal volume of phenol/chloroform (at a volume ratio of 1:1) was added to the final 2nd strand reaction mix, vortexed to form an emulsion and centrifuged on a microcentrifuge for 1 minute to separate the phases. The upper aqueous phase was retained and re-extracted with phenol/chloroform. 2 volumes of diethyl ether was added to the final aqueous phase and the mixture vortexed vigorously for several seconds. After centrifugation in a microcentrifuge for several seconds to separate the phases, the upper ether phase was discarded and the ether extraction repeated on the aqueous phase.

(b). Ethanol Precipitation

An equal volume of 4 M ammonium acetate was added to the final aqueous phase, followed by 2 volumes of absolute ethanol at -20 ^{o}C and the DNA was precipitated at -70 ^{o}C for 40 minutes.

The precipitate was warmed to room temperature to dissolve any unincorporated nucleotides and the DNA recovered by centrifugation in a microcentrifuge for 10 minutes at 4 ^{o}C.

The supernatant was removed and the pellet resuspended in 50 µl of TE buffer. 50 µl of 4 M ammonium acetate was added and the DNA was precipitated in 2 volumes of ethanol at -70 ^{o}C for 40 minutes. The precipitate was warmed to room temperature and centrifuged for 10 minutes in a microcentrifuge at 4 ^{o}C. The supernatant was removed and the pellet washed with 100 µl of 70 % ethanol then spun for 3 minutes in a microcentrifuge. The ethanol was removed and the pellet dried briefly *in vacuo*. The dry pellet was taken up in 20 µl of TE buffer and stored at -70 ^{o}C.

2.11.3 Estimation of Radioactivity Incorporated into Nucleic Acid.

A 1 µl sample of the DNA was spotted onto a Whatmann GF/C glass fibre disc and the disc allowed to dry.

A second 1 µl sample was added to a tube containing 100 µl of salmon sperm DNA solution (500 µg/µl in 20 mM EDTA). 5 ml of ice cold 10 % TCA was added and the mixture chilled on ice for 15 minutes. The precipitate was then collected by filtration through a second GF/C disc. This disc was washed 6 times with 5 ml of ice cold 10 % TCA, followed by one wash with 5 ml of 95 % ethanol. The disc was then allowed to dry.

Both discs were placed in scintillation vials and were counted in toluene based scintillation fluid. The count on the first disc gives a measure of the total radioactivity in a 1 µl sample and the count on the second gives a measure of the radioactivity incorporated into nucleic acid.

2.11.4 Calculations of Percentage Incorporation of Radioactivity and of Reaction Yields for cDNA Synthesis Reactions.

(a). Incorporation of Radioactivity into 1st strand cDNA

For the 1st strand reaction, the percentage incorporation of radioactivity is as follows:

$$\frac{\text{cpm of 2nd GF/C disc(radioactivity incorporated into nucleic acid)}}{\text{cpm of 1st GF/C disc(total radioactivity in the sample)}} \times 100$$

The amount of cDNA synthesised = (% incorporation X 140) ng.

Therefore, the percentage of mRNA transcribed

$$= \frac{\text{amount of 1st strand cDNA synthesised}}{\text{amount of input poly } (A^+) \text{ mRNA}} \times 100$$

(b). Incorporation of Radioactivity into 2nd Strand cDNA. For the second strand, the percentage incorporation must take into account the contribution of the radioactivity included in the 1st strand cDNA reaction mix. Since the final 2nd strand reaction volume (120 µl) is 6 X the 1st strand reaction volume

(20 µl), this can be accounted for as follows:

Percentage incorporation of radioactivity in the 2nd strand

$$= \frac{(D - B/6)}{(C - A/6)} \quad X\ 100$$

where:

A = total cpm in sample of 1st strand reaction mix;
B = cpm incorporated into nucleic acid in sample of 1st strand reaction mix;
C = total cpm in sample of 2nd strand reaction mix;
D = cpm incorporated into nucleic acid in sample of 2nd strand reaction mix.

The amount of 2nd strand cDNA synthesised

= (% incorporation of radioactivity in the 2nd strand X 140)ng

Therefore, the percentage of 2nd strand synthesised from 1st strand

$$= \frac{(\text{amount of 1st strand cDNA})}{(\text{amount of 2nd strand cDNA})} X\ 100$$

2.12 Ligation of cDNA into Plasmid pTZ18R

The optimum ligation conditions were worked out from the series of experiments detailed in Chapter 3. T4 DNA ligase was obtained from Gibco BRL and was used according to the manufacturer's instructions for blunt end ligation reactions. The buffer used was that supplied by the manufacturer.

A 3:1 (w:w) ratio of phosphatased *Sma* I cut pTZ18R to insert was used in the ligation reaction. cDNA used as insert was used in the ligation reaction immediately after synthesis. The reaction mix was as follows :

5 X ligase buffer	2 µl
Phosphatased pTZ18R	60 ng
cDNA	20 ng
T4 DNA ligase	2 units

This was made up to 10 µl with sterile distilled water and incubated overnight at room temperature.

After incubation, the complete ligation mix was transformed into competent *E.coli* NM522 cells, as described in section 2.6.

2.13 Preparation of Agarose Gels for Analysing Nucleic Acids.

2.13.1 Preparation of Agarose Gels for Analysing DNA.

Agarose gels were typically 0.7 % - 1.75 % agarose and were prepared as described in Maniatis *et al.*, 1982. An appropriate amount of agarose was dissolved in 30 ml of Tris-borate electrophoresis buffer, if the gel was a minigel, in 80 ml of Tris-borate electrophoresis buffer if the gel was a midigel, or in 150 ml of Tris-acetate electrophoresis buffer, if the gel was a large gel. This was heated to 100 ^{o}C on a Bunsen burner, then boiled gently for a few minutes until all the agarose had gone into solution. The agarose was rapidly cooled to 55 - 60 ^{o}C by swirling under running water, 0.5 µg/ml ethidium bromide was added, (unless the gel was a large gel, which was stained in 50 ml of 1 µg/ml ethidium bromide after electrophoresis), and the gel was then poured into a horizontal gel apparatus with a gel comb already in place. Once the agarose had set, the gel comb was removed, an appropriate volume of 1 X electrophoresis buffer (Tris-borate for minigels and midigels, Tris-acetate for large gels) was added, and the samples could then be loaded onto the gel.

Minigels were normally run for 30 minutes at 60 - 80 mA whereas large gels and midigels were run overnight at a constant voltage of 40 V.

2.13.2 Preparation of a Denaturing 1.2 % Agarose Gel with 2.2 M Formaldehyde.

This was prepared as described in Maniatis *et al.*, 1982.

1.2 % agarose was made up by dissolving 0.36 g of agarose in 24 ml of 1 X MOPS running buffer for minigels or by dissolving 2.4 g of agarose in 160 ml of 1 X MOPS running buffer for large gels. This was heated to 100 ^{o}C on a Bunsen burner, then boiled gently for a few minutes until all the agarose had gone into solution.

The agarose was cooled rapidly to approximately 55-60 ^{o}C and formaldehyde was added to a concentration of 2.2 M (6 ml of formaldehyde for minigels; 40 ml for

large gels).

The agarose was then poured into a horizontal gel apparatus with a perspex comb in position and allowed to set at room temperature.

Once the gel had set, the gel comb was carefully removed and 1 X MOPS running buffer added to the buffer reservoirs.

Samples were prepared as follows :
e.g. for a 20 µl sample:

RNA	up to 4.5 µl
5 X MOPS running buffer	2.0 µl
formaldehyde	3.5 µl
formamide	10.0 µl

This was made up to 20 µl with sterile distilled water.

2 µl of sterile loading buffer was added to each sample, and the samples were incubated at 60 ^{o}C for 15 minutes prior to being loaded on the gel.

Samples were then loaded onto separate sample wells on the gel and electrophoresed overnight at 60 V., if the gel was a large gel, or for 2 hours at 100 V., if the gel was a minigel. After electrophoresis, the gel was stained for 30 minutes in Acridine Orange (50 µg/ml in 10 mM sodium phosphate buffer, pH 7.0), then destained in 10 mM potassium phosphate, pH 6.3, overnight.

2.13.3 Preparation of a 1.4 % Alkaline Agarose Gel.

This was prepared as described in Maniatis *et al.*, 1982. 0.42 g of agarose was dissolved in 30 ml of 1 X alkaline gel preparation buffer, heated to 100 ^{o}C on a Bunsen burner, then boiled gently for a few minutes until all the agarose had gone into solution.

The agarose was rapidly cooled and poured into a minigel apparatus with a gel comb already in place. The gel was allowed to set at room temperature and the gel comb was then carefully removed. After it had set, the gel was covered to a depth of 5 mm with alkaline gel electrophoresis buffer and this was allowed to soak into the gel for at least 30 minutes.

Before loading the samples onto the gel, excess alkaline electrophoresis buffer

was removed to leave the gel covered to a depth of 1mm.

At this point, samples could be loaded onto the sample wells formed by the gel comb.

2.13.4 Alkaline Agarose Gel Electrophoretic Analysis of 1st and 2nd strand cDNA.

A sample representing 10000-30000 cpm incorporated into nucleic acid was placed in a 1.5 ml Eppendorf tube. 20 µl of salmon sperm DNA solution (100 µg/ml) was added and 1/3 of the combined volume of 1 M NaOH added. The mixture was incubated at 46 °C for 30 minutes. This step removes any remaining RNA from the samples.

After incubation, a volume of 1 M HCl equal to that of NaOH was added, followed by a volume of 1 M Tris-Cl, pH 8.0, equal to that of HCl.

An equal volume of phenol/chloroform was added and a phenol extraction carried out, followed by ethanol precipitation of the DNA, as described in section 2.11.2. After washing in 70 % ethanol, the dried DNA pellet was taken up in 20 µl of 1 X alkaline gel loading buffer, instead of TE buffer.

A 1.4 % alkaline agarose minigel was set up, as described in section 2.13.3, and samples were loaded onto the gel.

Samples were as follows:

1. 1st strand cDNA, prepared as described in above.
2. 2nd strand cDNA, prepared as described in above.
3. 5 µl of ^{32}P labelled λ/*Hin* dIII fragments
   + 5 µl of sterile distilled H_2O + 10 µl
   of 2 X alkaline gel loading buffer.

2.14 Radioactive Labelling of DNA probes

2.14.1 Nick Translation of DNA Probes

DNA probes were labelled by nick translation using (α-^{32}P)-dCTP and DNA polymerase I (Kornberg Polymerase) as described by Rigby *et al.*, 1977.

Nick Translation reactions were set up as follows :

DNA (100 - 250 ng)	1.0 - 5.0 µl
10 X NT buffer	5.0 µl
2 mM dATP	0.5 µl
2 mM dGTP	0.5 µl
2 mM dTTP	0.5 µl
(α-^{32}P)-dCTP	25 - 50 µCi
DNase I (0.1µg/ml)	0.5 µl
E.coli DNA Polymerase I	1.0 µl

This was made up to 50 µl with sterile distilled water and incubated at 15 °C for 90 minutes. The reaction was stopped by the addition of 5 µl of 0.2 M EDTA, pH 8.0. The mixture was made up to 100 µl and unincorporated nucleotides separated from labelled probe using the spun column procedure described by Maniatis *et al.*, 1982.

2.14.2 End Labelling of Oligonucleotides Using T4 Polynucleotide Kinase

Oligonucleotides were obtained from OSWEL DNA Service, University of Edinburgh.

100 ng of oligonucleotide was labelled with ^{32}P using T4 polynucleotide kinase as described in Maniatis *et al.*, 1982. The reaction was as follows :

oligonucleotide (100 ng)	1 µl
(γ-^{32}P)-ATP	20 µCi
10 X Kinase buffer	1 µl
T4 Polynucleotide Kinase	10 units

This was made up to 10 µl using sterile distilled water and incubated at 37 °C for 60 minutes. The reaction was stopped by the addition of 1 µl of 0.5 M EDTA, pH 8.0.

Unincorporated (γ-^{32}P)-ATP was separated from labelled oligo on a static Sephadex G25 column, as described in Maniatis *et al.*, 1982.

Hybridisations using labelled oligos were carried out at 37 °C.

2.14.3 Labelling Restriction Fragments with ^{32}P by Filling in the Cohesive Termini

DNA restriction fragments with recessed 3' termini were labelled using the Klenow fragment of *E.coli* DNA polymerase I according to the method described in Maniatis *et al.*, 1982. This method was used to label restriction fragments for use as

DNA size markers on agarose gels such as λ/*hin* dIII.

The DNA was digested to completion with restriction enzyme in the appropriate buffer, then the DNA was precipitated using isopropanol as described in section 2.1, and the precipitate resuspended in TE buffer.

1 µg of the digested DNA was labelled with an appropriate ^{32}P-dNTP in the following reaction :

DNA (e.g. λ/*Hin* dIII)	1 µg
10 X NT buffer	2 µl
2 mM dATP	1 µl
2 mM dGTP	1 µl
2 mM dTTP	1 µl
(α-^{32}P)-dCTP	2 µCi
Klenow fragment	1 unit

This was made up to 20 µl with sterile distilled water and incubated at room temperature for 30 minutes. The reaction was stopped by the addition of 1 µl of 0.5 M EDTA, and the labelled DNA was separated from unincorporated (α-^{32}P)-dCTP using the spun column method described in Maniatis *et al.*, 1982.

2.15 Transfer of Nucleic Acids to Nitrocellulose Filters

2.15.1 Transfer of DNA from Agarose Gels to Nitrocellulose Filters (Southern Blots).

DNA fragments, generated by restriction digests, were transferred from horizontal (1 - 1.75 %) agarose slab gels using the method of Southern, 1975.

The gel containing the DNA to be transferred was trimmed of excess agarose and then treated as follows :

- washed for 2 X 15 minutes in 0.25 M HCl to partially depurinate the DNA and facilitate transfer of large DNA fragments;
- washed for 2 X 15 minutes in 0.5 M NaOH, 1.5 M NaCl to denature the DNA;
- washed for 2 X 15 minutes in 1 M Tris-Cl, pH 7.5, 1.5 M NaCl to neutralise the gel;
- washed for 20 minutes in 10 X SSC to equilibrate the gel prior to transfer.

Nitrocellulose filters, cut slightly larger than the area of the gel, were floated on distilled water to "wet" the nitrocellulose, and then soaked in 10 X SSC for 20

minutes immediately prior to use.

All Southern blots were sandwich blots and the blotting components were prepared as described in Maniatis *et al.*, 1982. A 5 cm layer of paper towels was laid on a glass plate, followed by 4 sheets of dry Whatmann 3MM paper; 2 sheets of Whatmann 3MM paper soaked in 10 X SSC; a nitrocellulose filter; the gel containing the DNA fragments to be transferred; a second nitrocellulose filter; 2 sheets of 3MM paper soaked in 10 X SSC; 4 sheets of dry 3MM paper; a 5 cm layer of paper towels and finally a heavy glass plate to ensure even contact.

After transfer at room temperature for at least 90 minutes, each filter was marked in the top right hand corner for identification, and then baked *in vacuo* for 2 hours at 80 ^{o}C. Baked filters were sealed in thick polythene bags and stored at 4 ^{o}C. Immediately prior to use, the filters were washed for 15 minutes in 2 X SSC.

2.15.2 Transfer of RNA from a Denaturing Agarose Gel with Formaldehyde to Nitrocellulose Filters (Northern Blots).

RNA samples (total RNA and mRNA) were transferred from a large horizontal 1.2 % denaturing agarose gel with formaldehyde using the method of Southern, 1975, as modified by Thomas,1980.

The gel, containing the RNA samples to be transferred, was trimmed of excess agarose, then rinsed in sterile distilled water for several minutes to remove excess formaldehyde.

The gel was then placed on top of wicks of 3MM paper which overlapped a glass plate sitting in a tray of 20 X SSPE so that the wicks were in contact with a reservoir of 20 X SSPE.

A nitrocellulose filter, cut slightly larger than the area of the gel, was floated on distilled water for 10 minutes, to "wet" the nitrocellulose, then soaked in 20 X SSPE immediately prior to transfer.

The nitrocellulose filter was placed on top of the gel followed by several layers of 3MM paper, a 5 cm layer of paper towels, and finally a heavy glass plate to

ensure even contact.

After transfer at room temperature overnight, the nitrocellulose filter was marked in the top right hand corner for identification, prior to baking at 80 °C under vacuum for two hours. Baked filters were sealed in thick polythene bags and stored at 4 °C, prior to use.

Immediately before use, the filters were washed in 2 X SSC for 15 minutes.

2.15.3 Dot Blots

This method was a modification of the method for hybridisation of denatured RNA dotted onto nitrocellulose paper described by Thomas, 1983.

A strip of nitrocellulose paper was cut and "wet" by floating on RNase-free distilled water for at least 5 minutes. The nitrocellulose was then equilibrated with 20 X SSC for 15-30 minutes, after which the nitrocellulose was air-dried.

Samples were prepared in 4 µl volumes and dotted onto the dried nitrocellulose in 1 µl fractions, each fraction being dried with a hairdrier prior to the addition of the next fraction.

After all the samples had been dotted onto the nitrocellulose and dried, the filter was baked at 80 °C for 2 hours under vacuum. The baked filters were then sealed in plastic and either hybridised immediately or stored at 4 °C.

2.15.4 Colony Blots

Plasmid DNA from *E.coli* colonies was blotted onto nitrocellulose, using a modification of the method described by Hanahan and Meselson, 1980. Transformed *E.coli* colonies were transferred to nitrocellulose and stored at -70 °C as colony "sandwiches", as described in section 2.7, prior to use.

The nitrocellulose disc "sandwich" was removed from the -70 °C freezer and allowed to thaw at room temperature. The discs were separated and a second replica was made by keying a fresh sterile nitrocellulose disc to the master template disc. This was stored at -70 °C until use.

The first replica was placed colony side up on an LB agar plate and incubated at

37 °C until distinct colonies of 2 - 3 mm in diameter had developed. The nitrocellulose disc was transferred, colony side up, to a dish containing several sheets of 3MM paper, barely saturated in 0.5 M NaOH for 5 minutes. This process was repeated, then the disc was transferred to several sheets of 3MM paper barely saturated with 1 M Tris-Cl, pH 8.0 for 2 X 5 minutes; then to several sheets of 3MM paper barely saturated with 1 M Tris-Cl, pH 8.0, 1.5 M NaCl for 2 X 5 minutes. The disc was allowed to dry in air and was then baked *in vacuo* for 2 hours at 80 °C.

Immediately prior to use, the filter was placed in 2 X SET for 15 minutes.

2.15.5 Colony Hybridisation on Whatmann 541 Filters

This was a modification of the method described by Gergen *et al.*, 1979.

Transformed *E.coli* cells were plated on LB agar plates and incubated at 37 °C until the colonies reached 1 - 2 mm in diameter.

A sterile Whatmann 541 disc was placed on the surface of the agar plate and the plate was incubated at 37 °C for 2 hours. The paper was carefully removed from the surface of the agar plate and placed colony side up in the following :

- 2 X 5 minute washes in 0.5 M NaOH;
- 2 X 5 minute washes in 0.5 M Tris-Cl, pH 8.0;
- 2 X 5 minute washes in 2 X SSC;
- 1 X 2 minute wash in 95 % ethanol.

The disc was air-dried, sealed in thick plastic, and either stored at 4 °C or hybridised immediately with radioactive probe for 4 hours at 65 °C in 5 X SSC, 250 µg/ml salmon sperm DNA.

After hybridisation, the disc was washed twice in 500 ml of 2 X SSC at room temperature, dried in air and autoradiographed at -70 °C.

2.16 Hybridisation of Filters Using Radioactive Probes

2.16.1 Hybridisation of Southern Blots with Radioactive Probe

Nitrocellulose filters prepared by Southern blotting as described in section 2.15.1 were prehybridised for 4 hours at 65 °C in the following solution in sealed

polythene bags :

10 X SSC	10.0 ml
50 X Denhardt's solution	2.0 ml
10 % SDS	0.2 ml
0.5 M EDTA, pH 8.0	0.1 ml
Salmon Sperm DNA(5 mg/ml)	0.4 ml
Dextran Sulphate	2.0 g

This was made up to 20 ml with sterile distilled water and heated to 65 ^{o}C before adding to the filters.

Radioactive probes were prepared as described in sections 2.14.1 and 2.14.2 and denatured by heating to 100 ^{o}C for 10 minutes before adding to the prehybridisation mix in the polythene bags. The probe was thoroughly mixed in and air bubbles were expelled before resealing the bags and allowing hybridisation to proceed overnight at 65 ^{o}C in a shaking water bath.

After hybridisation, the filters were removed and washed successively in the following :

- 5 X SSC; 0.1 % SDS for 30 minutes at 65 ^{o}C.
- 2 X SSC; 0.1 % SDS for 60 minutes at 65 ^{o}C.
- 0.2 X SSC; 0.1 % SDS for 30 minutes at 65 ^{o}C.

The filters were blotted dry on Whatmann 3MM paper and sealed in fine polythene bags.

Each filter was autoradiographed by exposure to either Kodak Ortho G or Fuji-RX X-ray film in Kodak X-omatic cassettes using Lanex regular intensifier screens at -70 ^{o}C.

2.16.2 Hybridisation of Northern Blots and Dot Blots with Radioactive Probes

Nitrocellulose filters prepared by Northern blotting, as described in section 2.15.2, or by dot blotting, prepared as described in section 2.15.3, were prehybridised for 4 hours in the following solution (20 ml for Northern blots; 10 ml for dot blots) in sealed polythene bags :

4 X SSC
50 % formamide
6 X Denhardt's solution
20 µg/ml Salmon Sperm DNA

0.1 % SDS

This was heated to 42 °C before adding to the filters.

Radioactive probes were prepared as described in section 2.14.1 and were denatured by heating to 100 °C for 10 minutes before adding to the prehybridisation mix in the polythene bags. Hybridisation was allowed to proceed at 42 °C for 48 hours.

After hybridisation, the filters were washed succesively as follows :

- 2 X SSC; 0.5 % SDS for 20 minutes at room temperature.
- 0.1 X SSC; 0.1 % SDS for 120 minutes at room temperature.
- 0.1 X SSC; 0.1 % SDS for 30 minutes at 65 °C.

The filters were blotted dry on Whatmann 3MM paper and sealed in fine polythene bags.

Each filter was autoradiographed as described for Southern blots.

2.16.3 Hybridisation of Colony Blots with Radioactive Probe

Colony blots on nitrocellulose filter discs, prepared as described in section 2.15.4 were prehybridised in the following solution in sealed polythene bags for 1 - 2 hours at 65 °C :

20 X SET	12.5 ml
50 X Denhardt's solution	1.0 ml
10 % SDS	0.5 ml

This was made up to 50 ml with sterile distilled water.

Radioactive probe, prepared as described in section 2.14.1 or 2.14.2, was denatured by heating to 100 °C before adding to the prehybridisation mix in the polythene bags. Hybridisation was allowed to proceed overnight at 65 °C.

After hybridisation, the filters were washed successively as follows :

- 2 X SET; 0.5 % SDS for 30 minutes at 65 °C.
- 1 X SET; 0.5 % SDS for 60 minutes at 65 °C.
- 0.2 X SET; 0.5 % SDS for 30 minutes at 65 °C.

The filters were blotted dry on Whatmann 3MM paper and autoradiographed as

described for Southern blots.

Chapter 3 : The Construction of cDNA Libraries from Neurospora crassa mRNA

Chapter 3

The Construction of cDNA Libraries from *Neurospora crassa mRNA*

3.1 Introduction

3.1.1 Synthesis of cDNA

As discussed in the introduction to this thesis, the system of choice for studying the behaviour of the *Neurospora crassa am* gene in response to treatment with frameshift mutagens while resident in *E.coli* involved the production of the gene product. Since the *am* gene contains 2 introns, it is not possible to express the gene itself in *E.coli*.

A powerful technique which can be used to express eukaryotic genes in prokaryotic cells is to make a cDNA copy of the gene of interest and use this as the template for expression in the prokaryote (see Maniatis *et al.*, 1982). There are many methods available for the production of cDNA libraries. However, the one most commonly used is that described by Gubler and Hoffman, 1983, which uses reverse transcriptase to produce a DNA strand which is complementary to the mRNA. The mRNA in the cDNA/mRNA hybrid is then 'nicked' using RNase H and the second cDNA strand transcribed using *E.coli* DNA polymerase I. In this system, the first strand of the cDNA is used as template and the short RNA molecules which result from the RNase H treatment act as primers for the second strand. Following synthesis, the double stranded cDNA molecules can be made blunt ended by using the 3' exonuclease activity of T4 DNA polymerase to remove 3' overhangs and the cDNA can either be directly ligated into a blunt-ended vector, or linkers or adaptors carrying a cohesive end restriction site can be ligated to the cDNA molecule to allow it to be ligated into the same cohesive end restriction site on the vector of choice.

3.1.2 Choice of Vector

When deciding upon the most appropriate vector for the construction of a cDNA library from *Neurospora crassa* with the express aim of isolating a cDNA clone for the *am* gene, several important considerations must be taken into account.

These are:

1. Although a representative cDNA library would be of interest, the most important feature required of the library is representation of a full length cDNA for the *am* gene.
2. The *am* sequence should be easily isolated from the library for characterisation.
3. Since the mutagenesis assay envisaged for the *am* cDNA requires that the cDNA can be expressed in *E.coli*, some sort of *E.coli* expression system is required.
4. PA340, the *E.coli* strain auxotrophic for glutamate, which will be used as a host for the mutagenesis assay, is resistant to lambda bacteriophage and is a poor transformation host for plasmids.

The vector of choice for cDNA libraries, under normal circumstances, is lambda bacteriophage since the use of λ vectors allows the formation of larger and potentially more representative libraries than is possible using a plasmid vector. However, a major disadvantage to using λ vectors for cDNA cloning is that, following the initial cloning and identification of the cDNA of interest, more subsequent manipulation of the insert DNA is required. For example, once a cDNA clone of interest has been identified in a lambda library, the insert cDNA must then be excised from the vector DNA, purified and subcloned into a plasmid vector before further manipulation is possible. Part of the reason for this subcloning step is that the insert DNA can be more easily characterised by restriction mapping when it is present in a small plasmid molecule, than would be possible in λ, which is typically around 40 kB in size. Since the average cDNA insert is unlikely to lie outside of the 0.5-3.0 kB size range, the possibility of finding unique restriction sites within the insert DNA, which are not represented in the vector DNA, is very much increased when that vector is a 3-5 kB plasmid DNA molecule than it would be if the insert DNA was contained within a 40 kB λ DNA molecule.

Although the ease of manipulation of the insert DNA, subsequent to subcloning in a plasmid vector, provides a powerful argument for transferring the insert from λ to a plasmid, there would be an even more important reason for transferring the cDNA insert into a plasmid vector for the purposes of this study. The *E.coli* glutamate auxotroph, PA340, which was to be used in the *am* cDNA expression and mutagenesis studies, is resistant to infection by λ bacteriophage

since it carries a mutation in the *malA1* gene and so the use of a λ vector in this strain would not be appropriate. PA340 is also a poor transformation host for plasmids so it could also be argued that these vectors would be inappropriate for this strain. However, this problem can be overcome by first identifying the plasmids containing the cDNA of interest in a different *E.coli* strain, one which is a good transformation host, isolating these plasmids after identification, and then using them to transform PA340.

Another consideration which had to be taken into account when choosing a vector for this study was the requirement for expression of the insert DNA after transfer to PA340. In the Introduction to this thesis, it was mentioned that the *am* sequence does not contain transcription or translation signals which would be recognisable to *E.coli*. Therefore, these would have to be provided by the vector which mediates the transfer of the *am* cDNA into PA340. The easiest way to provide these signals is to fuse the *am* sequence to part of the coding sequence of an *E.coli* gene and express it as a fusion protein. By placing the *am* cDNA close to the beginning of the translated portion of the *E.coli* gene, interference in the *am* gene product function by the *E.coli* protein can be minimised. By choosing an *E.coli* gene which shows inducible expression and which can, therefore, be tightly controlled, as the gene to which the *am* cDNA would be fused, expression of the *am* gene product could also be controlled. Although it is not essential that the vector used in the construction of the cDNA library also provides the appropriate expression signals, it would be advantageous to choose a vector which could act as an expression vector at later stages in the study as this would help to minimise the amount of subsequent manipulations and alleviate the necessity to subclone.

Although the use of a plasmid vector would lead to a potentially smaller and less representative cDNA library, the most important consideration, for the purposes of this study, was that the *am* cDNA be represented within the library. Therefore a library of lower titre need not cause any problems, provided that the *am* cDNA sequence was represented within the library. Careful consideration of the advantages and disadvantages offered by plasmids versus lambda bacteriophage as the vector for the cDNA library led to the decision that, for the

purposes of this study, a plasmid vector offered more advantages than a lambda vector, although at the cost of a potentially smaller cDNA library.

The use of a plasmid vector to construct the cDNA library offered the following advantages:

1. If chosen correctly, the same vector could be used throughout the study. By choosing a plasmid which contained a cloning site within an inducible *E.coli* gene, and cloning the cDNA library into that site, a single vector could be used for cloning the cDNA, identifying the *am* gene, and for expression of the gene product after transfer to PA340. This would eliminate the requirement to subclone the *am* cDNA, after identification.
2. After identification of potential *am* cDNA clones, the recombinant plasmids containing the cDNA could be easily purified and the DNA mapped using restriction enzymes to confirm the presence of *am* cDNA.
3. Several plasmid vectors offer a system for identifying recombinant plasmids based on the ability of β-galactosidase, the gene product of the *lacZ* gene of *E.coli*, to cleave the chromogenic substrate, X-gal, and produce a blue colour. Ligation of insert DNA into the *lacZ* gene inactivates its ability to produce functional β-galactosidase and so recombinant colonies appear white. This allows identification of colonies harbouring plasmids which do not contain inserts so that the level of 'background' non-recombinant colonies can be established. The *lacZ* gene of *E.coli* has the additional advantage that it is controlled by an inducible *E.coli* promoter, so this system would also provide the possibility for inducible *E.coli/am* fusion protein expression.
4. The presence of polylinker sequences, containing a number of unique restriction sites, on some plasmids allows a choice of cloning strategies and aids mapping and isolation of the cDNA insert after identification.

The ideal situation would have been to clone the *N.crassa* cDNA library into a plasmid under the control of an inducible *E.coli* promoter, transform the entire library into PA340 and identify *am* cDNA clones by their ability to restore glutamate prototrophy to the glutamate auxotroph, PA340, upon induction. Indeed, a method similar to this was used by Moye *et al.*, 1985, to obtain the *GDH1* gene from *S.cerevisiae*.

However, there were several obstacles to using this approach to find the *am* cDNA clone from *N.crassa*. Firstly, PA340 is a poor transformation host, and, secondly, it was not known whether or not the *am* gene product was capable of complementing glutamate auxotrophy in *E.coli*. Some early attempts to show complementation of *gdh*$^{-}$ strains of *E.coli* using the *am* gene by Dr. Jane

Kinnaird at Edinburgh had proved unsuccessful (personal communication) but these studies were attempted before the *am* DNA sequence was known and there was no way to predict the presence of the two small introns within the gene. The use of a cDNA clone to the *am* gene to complement glutamate auxotrophy in *E.coli* had not previously been attempted.

A third obstacle to directly screening the cDNA clones in PA340 lay in the nature of glutamate auxotrophy in *E.coli*. As stated in the Introduction to this thesis, glutamate auxotrophy in *E.coli* requires that two genes, *gdhA*, encoding glutamate dehydrogenase, and *gltB*, encoding glutamate synthase, are inactive. In addition to the lack of information about the behaviour of the *am* gene product in a glutamate auxotroph of *E.coli*, there was no available information about the behaviour of the *N.crassa* glutamate synthase in a glutamate auxotroph of *E.coli*. Since *E.coli* expression studies have shown that glutamate synthase is more efficient at restoring glutamate prototrophy than glutamate dehydrogenase (Berberich *et al.*, 1972; Pahel *et al.*,1978), this opened up a possibility not only for the generation of false positives, but also for the preferential selection of glutamate synthase clones.

For these reasons, the following approach was considered the most appropriate for this study.

After insertion of the cDNA into a plasmid vector under the control of an inducible *E.coli* promoter, the library could be transformed into an *E.coli* strain which would allow high efficiency transformation and therefore optimise the library size. cDNA clones containing the *am* sequence could be identified by their ability to hybridise to the *am* genomic clone in colony hybridisation experiments. Once identified, the recombinant plasmids containing the *am* cDNA could be mapped using restriction enzymes to confirm their identity as *am* cDNA clones and to determine the orientation of the inserts, with respect to the promoter. *am* cDNA clones which were in the correct orientation for expression could then be used to transform PA340 for expression studies. Using this approach, the fact that PA340 is a poor transformation host does not present a problem since large numbers of transformants are unnecessary because any PA340 transformants obtained would contain the plasmid of interest.

The plasmid used would have to provide both the transcription and the translation signals for expression of the *am* cDNA in *E.coli*. A common method for obtaining expression of cDNA sequences in *E.coli* is to fuse the cDNA sequence to the coding sequence of the *lacZ* gene of *E.coli*. Indeed, this provides the basis for cDNA clone identification using some vector systems. By cloning the cDNA into a unique restriction site at the 3' end of the *lacZ* coding sequence in vectors such as λgt 11 (Young and Davis, 1983), and the pEX series of plasmids (Stanley and Luzio,1984), and then inducing expression of the fusion products using IPTG, cDNA clones which express the cDNA of interest can be identified using antibodies specific to the cDNA product. However, the fusion protein, although still recognisable using antibodies to the foreign gene product, very rarely retains the ability to operate as a functional protein. Since NADP-GDH from *N.crassa* may not be active in *E.coli* in its native conformation, it was decided that limiting the amount of *E.coli* derived protein which was fused to the *am* cDNA was desirable.

There are several plasmid vectors which contain only the *lacZ'* gene of *E.coli*, which encodes only the most 5' end of the *lacZ* gene. Plasmid vectors based on the pUC range of vectors (Vieira and Messing, 1982) have been engineered to contain a polylinker sequence within the N-terminal coding sequence of the *lacZ'* gene. This means that expression of inserts fused to the *lacZ'* coding sequence leads to the insertion of only a few additional *E.coli* derived amino acids at the N-terminus of the translation product. Expressing the *am* cDNA as a fusion product requires the cDNA not only to be in the correct orientation for expression but also to be in the same frame as the *lacZ'* coding sequence. This means that only one in three *am* cDNA clones identified as being in the correct orientation would also be in the correct reading frame for expression. However, since the sequence of the genomic *am* clone was known, it should be possible to use that information to manipulate the cDNA sequence and bring it into frame with the *lacZ'* sequence of the plasmid, if necessary.

The pUC based plasmids have the following features which make them ideal for the purposes of this study:

1. The plasmids confer ampicillin resistance on the *E.coli* host so *E.coli* cells

containing plasmids can be selected.

2. The presence of a polylinker region, within the plasmid DNA, which contains several unique restriction sites, allows a choice of cloning sites and strategies.
3. Since the polylinker is contained within the *lacZ'* coding sequence, recombinant colonies can be easily identified by their inability to metabolise the chromagenic substrate, X-gal.
4. The position of the polylinker in the *lacZ'* gene means that the *lacZ'* transcription and translation signals can be used to express the *am* cDNA as a fusion protein. In addition, by selecting a cloning site which is central within the polylinker, unique sites 5' of the cloning site would be available for further manipulation e.g. to correct reading frame differences between the *lacZ'* coding sequence and the cDNA sequence.
5. pUC plasmid pairs are available which contain the polylinkers in opposite orientations, e.g. pUC 19 has the same polylinker as pUC 18 but the pUC 19 polylinker is in the opposite orientation from that of pUC 18. This means that clones which are identified as containing the correct insert but which contain that insert in the wrong orientation in pUC 18 can be placed in the correct orientation by transferring the entire polylinker sequence, including the insert, into pUC 19.
6. The expression of *lacZ'*, and therefore the fusion protein, can be tightly controlled since the presence of an inducer is required for expression.
7. The plasmids are high copy number plasmids, therefore chloramphenicol amplification would be unneccesary, prior to screening the cDNA clones. The high copy number may also be useful if *N.crassa* NADP-GDH is not as efficient in *E.coli* as it is in *N.crassa*.

In 1986, Mead *et al.* introduced the pTZ18R/U and pTZ19R/U phagemid vectors. These vectors are based on the pUC 18 and pUC 19 plasmids. In addition to the features listed for the pUC based plasmids, the pTZ plasmids contain an F1 origin of replication, which allows the production of single stranded phagemid DNA after superinfection with an M13-based helper phage, and a T7 promoter at the 5' end of the polylinker, which allows transcription of the insert *in vitro* using T7 RNA polymerase. Both of these features may prove useful in future studies on the *am* cDNA. The ability to rescue a single strand of the plasmid DNA facilitates *in vitro* mutagenesis and preparation of DNA for single stranded sequencing. Since any mutants generated later in the study would have to be analysed by DNA sequencing, this would be an extremely useful feature to have available.

The ability to produce transcripts of the cDNA using T7 RNA polymerase could prove useful as the transcripts could be translated *in vitro* and the translation

products tested for GDH activity in cell-free assays. This may be necessary if *N.crassa* NADP-GDH is not capable of complementing glutamate auxotrophy in *E.coli* as it would be essential to show that the *am* cDNA clone was capable of producing active NADP-GDH.

After some initial experiments using pUC 8 as the cloning vector, it was decided that the additional advantages offered by the pTZ plasmids made them more suitable vectors for use in this study. pTZ18R was chosen as the vector for the cDNA library. The pTZ18 vectors have the same polylinker sequence as pUC 18 and the pTZ19 vectors have the same polylinker sequence as pUC 19. The terms R and U refer to which strand of the DNA is rescued after superinfection with helper phage, the strand designated 'R' can be sequenced using the M13 'R'everse primer as primer, whereas the strand designated 'U' can be sequenced using the M13 'U'niversal primer as primer.

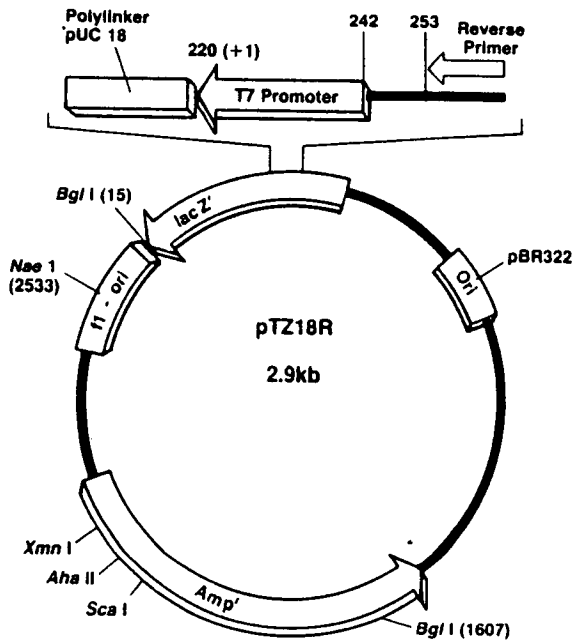
3.1.3 The Cloning Strategy Used

From the map of pTZ18R shown in figure 3.1, it is clear that a number of restriction sites are available within the polylinker for cloning. This presented a number of options for cloning the cDNA into this plasmid. The restriction site most commonly used for cDNA cloning is *Eco* RI. However, the position of the *Eco* RI cloning site on pTZ18R as the most 5' of the cloning sites on the polylinker, means that there are no sites 5' of it which can be used if enzymic manipulation is necessary to bring the cDNA coding sequence into the correct reading frame for expression. Therefore it was necessary to use a restriction site which was more central within the polylinker.

After careful consideration of the other cloning sites within the polylinker, it was decided that, for the purposes of this study, the most useful cloning strategy would be to ligate the blunt-ended cDNA into the *Sma* I site, immediately after cDNA synthesis. Although blunt end ligation is less efficient than sticky end ligations, the *Sma* I site was chosen because it offered the following advantages:

- The *Sma* I site is central within the polylinker which means that there are unique restriction sites 5' of the *Sma* I site which can be utilised if further manipulation is required to bring the cDNA coding sequence into the same frame as the *lacZ'* coding sequence.

FIGURE 3.1



Polylinker Sequence

LacZ' start codon

Eco RI *Sac* I *Sma* I *Kpn* I *Bam* HI *Xba* I *Sal* I *Pst* I *Sph* I *Hin* dIII

ATG ACC ATG ATT ACG AAT TTA ATA CGA CTC ACT ATA GGG AAT TCG AGC TCG GTA CCC GGG GAT CCT CTA GAG TCG ACC TGC AGG CAT GCA AGC TTG GCA CTG GCC GTC GTT TTA

T7 promoter

- The central position of the *Sma* I site also allows the same strategy to be used if the cDNA is inserted into pTZ18R in the wrong orientation for expression since, after transfer of the polylinker to one of the pTZ19 phagemids to change the orientation of the insert, there would still be a number of unique restriction sites 5' of the *Sma* I site which could be used to manipulate the reading frame of the cDNA.
- The amount of manipulation of the cDNA prior to cloning is reduced if the cDNA is blunt-end ligated into the *Sma* I site compared to ligating it into a sticky end restriction site, which requires methylation of the cDNA, addition of linkers, restriction digestion and separation of the cDNA from the digested linkers prior to ligation of the cDNA into the vector. Since each manipulation step can result in some loss of cDNA, it is desirable to keep the amount of manipulation to a minimum.

There were, however, two disadvantages to using the *Sma* I site as the cloning site. Firstly, after cloning into the *Sma* I site, the site is not normally recreated so could not be used to isolate the insert DNA; and, secondly, blunt end ligations are not as efficient as sticky end ligations, therefore the efficiency of ligation would have to be optimised prior to using the *Sma* I site as a cloning site. Optimisation of blunt end ligations for using the *Sma* I site as a cloning site is discussed further in section 3.6.4 of this chapter.

The fact that the *Sma* I site is not recreated does not normally cause problem because its central position within the polylinker allows the use of other restriction sites to recover the insert. Indeed, since the sequence of the *am* genomic clone is known, unique restriction sites on either side of the *Sma* I site of pTZ18R can be identified which are not represented within the *am* sequence.

For these reasons, the cloning strategy decided upon was to synthesise blunt ended cDNA fragments, ligate these fragments into the *Sma* I site of pTZ18R and then transform the resulting cDNA library into an *E.coli* strain using a high efficiency transformation protocol in order to maximise the size of the cDNA library.

The *E.coli* strain best suited for use with the pTZ plasmids is NM522. This strain carries the $lacI^q$ mutation which means that the strain overproduces the inhibitor for *lac* expression. This allows strict control of expression of the *lacZ'* gene. This strain also allows blue/white colour selection for non-recombinants

and can be superinfected with helper phage for preparation of single stranded DNA from the plasmid for sequencing.

After transformation of the cDNA library into NM522, the cells can be grown in the presence of ampicillin so that only those cells containing plasmids can grow. Expression of the *lacZ'* gene can be induced using IPTG and the proportion of non-recombinant colonies in the library determined by their ability to cleave the chromogenic substrate, X-gal, and appear blue in colour.

cDNA clones which carry the cDNA to the *am* gene can be identified by their ability to hybridise to the *am* genomic clone in colony hybridisation experiments. The orientation of the cDNA inserts with respect to the *lac* promoter can be determined by restriction analysis and potential *am* cDNA clones which are in the correct orientation for expression can be transformed into PA340 and tested for their ability to complement glutamate auxotrophy in the *E.coli* strain. *am* cDNA clones which contain the cDNA in the wrong orientation for expression can be transferred to pTZ19, which contains the polylinker in the opposite orientation to that of pTZ18, prior to transformation into PA340 to test for their ability to complement glutamate auxotrophy. Although a more ordered approach to expression of *am* cDNA in PA340, i.e. one which determined the frame of the *am* cDNA first and only transformed *am* cDNA clones in which the *am* coding region was in frame with respect to the *lacZ'* coding region into PA340, would be desirable, it was felt that the 'shotgun' approach adopted would provide information more quickly. Ensuring that only 'in frame' *am* clones were transformed into PA340 for complementation studies would involve determining the 5' sequence of each individual *am* clone so that the frame could be identified. Even then, there would be no guarantee that the cDNA would be expressed. It was also unlikely that an *am* cDNA clone would be produced which would be in frame with the *lacZ'* coding region and active for the following reasons:

- only half of the cDNA clones would be in the correct orientation for expression, although the other half could be transferred to pTZ19 to change their orientation;
- of those which were in the correct orientation, only one in three would be in the same reading frame as the *lacZ'* coding sequence;
- not all the cDNA clones would be full length so some may be missing

sequence information at their 5' or 3' end which may interfere with their ability to produce active NADP-GDH;
- clones which were full length would carry additional information at their 5' end, derived from the non-translated area of the mRNA, which will contribute additional amino acids to the N-terminal portion of the fusion protein and may interfere with the activity of the protein.

It should, therefore, be easier to test whether a particular *am* cDNA clone is capable of complementing glutamate auxotrophy by transforming the cDNA into PA340, inducing expression of the fusion protein and looking for the ability of that fusion protein to restore glutamate prototrophy. If no *am* cDNA clones which were capable of restoring prototrophy were found, then the frame of the cDNA coding sequence could be investigated and the DNA sequence manipulated to produce clones in which the *am* coding sequence was in the correct reading frame for expression.

3.1.4 Increasing the Representation of *am* cDNA clones within a *Neurospora crassa* cDNA Library

Since *am* mRNA is not an abundant mRNA species, it was unlikely to be highly represented within a *Neurospora crassa* cDNA library. It was therefore important to find some way of enriching the library for *am* cDNA clones. Initially, size selection of the cDNA prior to cloning was used as a method to enrich the library for *am* cDNA sequences but cDNA prepared in this way proved problematic at later stages of the cloning procedure, and resulted in low titre libraries. This is discussed in more detail in the results section of this chapter. Since size selection of the cDNA after synthesis yielded low titre libraries, it was decided that the best way to enrich the library for *am* sequence would be to select the mRNA prior to cDNA synthesis. There were two possible ways to do this. The mRNA could be selected on the basis of size so that only mRNAs of a similar size to *am* mRNA would be used for cDNA synthesis, or *am* mRNA could be specifically selected by its ability to hybridise to the genomic clone for the *am* gene. The latter method was thought to provide the best way to increase the representation of *am* cDNA clones within the cDNA library.

Hybridisation selection procedures which select a particular mRNA species by its ability to hybridise to a specific genomic or cDNA clone are normally used to prepare mRNA for *in vitro* translation and had not been previously reported as a

way to specifically select mRNA for conversion to cDNA. Although no technical difficulties were envisaged in using hybridisation selection as a way to prepare *am* mRNA for conversion to cDNA, it seemed sensible to prepare a cDNA library from non-selected *Neurospora crassa* mRNA as well as from the hybridisation selected *am* mRNA and to treat the cDNA libraries derived from *am*-selected and non-selected mRNA as separate libraries throughout the screening procedure. In this way, if any problems were encountered with the hybridisation selected *am* mRNA, there would be a library available which had been synthesised from total *N.crassa* mRNA, and had not been adapted in any way. The total cDNA library from *N.crassa* would also act as a control library for the *am* selected library.

3.2 Results and Discussion

3.2.1 Isolation of RNA from *N.crassa*

Total RNA was isolated from *Neurospora crassa* using the method detailed in section 2.8 of the Materials and Methods. The amount of RNA isolated was estimated from its $O.D._{260nm}$ using a conversion factor of 1 $O.D._{260nm}$ unit = 40µg/ml RNA. The yields of total RNA from *N.crassa* dried mycelia using this method are given in table 3.1. The quality of the RNA was determined by the presence of undegraded 18S and 28S ribosomal RNA bands on a 1.2 % agarose gel containing formaldehyde, prepared as described in section 2.13.2 of the Materials and Methods. The RNA on the gel was visualised by staining the gel with 50 µg/ml Acridine Orange solution for 15-60 minutes, followed by destaining with 20mM KPO_4, pH 6.3 for several hours until the RNA was clearly visible. A typical RNA gel showing the presence of undegraded ribosomal RNA bands is shown in figure 3.2.

3.2.2 Preparation of mRNA

Having established that the total RNA preparation showed no signs of degradation, the poly(A^+) mRNA was isolated on an oligo dT cellulose column, prepared as described in section 2.9 of the Materials and Methods.

Typically 8 mg of total RNA was loaded onto the oligo dT cellulose column.

Table 3.1

Yields of Total RNA from Dried N.crassa Mycelia and of mRNA after Oligo dT fractionation of Total RNA.

Starting Material (dried mycelial cells) (g.)	Amount of Total RNA isolated (mg RNA/g mycelia)	mRNA preparation	Amount of Total RNA loaded on to oligo dT column (mg)	% Yield of mRNA after oligo dT fractionation	mRNA used for cDNA synthesis
0.5 g	21.00 mg/g	1	10.0 mg	0.30 %	No
3.0 g	10.17 mg/g	2	8.0 mg	†0.60 %	Yes[A]
4.1 g	16.57 mg/g	3	8.0 mg	†0.56 %	No
		4	8.1 mg	0.39 %	No
2.4 g	22.46 mg/g	5	8.0 mg	0.80 %	No
		6	8.0 mg	0.50 %	Yes[B]
		7	8.0 mg	0.69 %	No
		8	8.0 mg	2.60 %	No
		9	8.0 mg	2.00 %	Yes[C]

† In these samples no peak was resolved initially by oligo dT fractionation. The RNA in the eluted fraction was therefore precipitated using ethanol, resuspended in binding buffer and subjected to a second oligo dT fractionation, which produced the results shown.
A,B,C refer to designations given to the mRNA preparations used for cDNA synthesis in table 3.2.

After elution of the poly(A^+) mRNA as described in section 2.9 of the Materials and Methods, the quantity of mRNA recovered was estimated from its $O.D_{260nm}$. Typical yields of poly(A^+) mRNA isolated from total *N.crassa* RNA after oligo dT fractionation are shown in table 3.1.

Initial attempts at poly(A^+) mRNA isolation using this method showed poor yields, with recovery typically less than 1.0 % of the total RNA which was loaded onto the column (compare mRNA preparations 1-7 on table 3.1). In addition, several early preparations did not show any discernible $O.D_{260nm}$ peak when the eluted fractions were checked for nucleic acid content. This problem was eventually traced to the batch of oligo dT cellulose used for the columns and was resolved by using a fresh batch of oligo dT cellulose from a different supplier (Boehringer Mannheim).

By using 0.45g of this oligo dT cellulose, which had a binding capacity of 100 A_{260nm} units per gram of dry powder, the yield of poly(A^+) RNA recovered was increased to 2 % or more of the total RNA loaded on the column (mRNA preparations 8 & 9, table 3.1.) mRNA eluted from the oligo dT cellulose columns was checked for degradation by electrophoresis on a 1.2 % agarose gel containing formaldehyde. As can be seen in figure 3.2, the 18S and 28S ribosomal bands are still evident in the mRNA preparation. Repeating the oligo dT cellulose fractionation on the poly(A^+) RNA did not show any significant loss of this ribosomal RNA. A possible explanation for this is the presence of polyadenylated stretches within the ribosomal RNA which may also bind to the oligo dT and therefore remain in the mRNA preparation. However, the presence of 18S and 28S ribosomal RNA in the preparation did provide a way to identify degradation in the mRNA preparation since degradation would lead to loss of the rRNA bands. The presence of rRNA bands in the mRNA lanes of figure 3.2 would seem to indicate that the RNA was free from degradation. To confirm that this was the case, a Northern blot of the RNA was performed according to the method in section 2.15.2 of the Materials and Methods, and this was probed using a genomic clone of the *am* gene from *Neurospora crassa*, which had been labelled with ^{32}P by nick translation according to the method described in section 2.14.1, to check whether the mRNA for the *am* gene was represented

within the mRNA as a single band. which would infer that it had not suffered any degradation.

The result of a typical Northern blot is shown in figure 3.3. From this figure, it can be seen quite clearly that there is a single band, of approximately 1.8 kB in size, evident in both the mRNA and total RNA fractions. The size of the mRNA calculated from the Northern blot agrees with the expected size for *N.crassa am* mRNA so this would indicate that the mRNA for the *am* gene was present intact within the poly(A^+) mRNA preparation. The presence of a single undegraded band which was homologous to the *am* genomic clone on a Northern blot was the criterion used for determining which mRNA preparations to use as templates for cDNA synthesis. Table 3.1 indicates which mRNA preparations were used for cDNA synthesis.

3.2.3 Hybridisation Selection of *am* mRNA

Prior to cDNA synthesis, some of the *N.crassa* mRNA was enriched for the *am* sequence by using the ability of the *am* mRNA to hybridise to the genomic *am* clone. The *am* genomic DNA clone was immobilised on a solid support, in this case, nylon Genescreen plus membrane, and *am* mRNA selected from a pool of other mRNAs by hybridisation to the *am* genomic clone on the membrane. mRNAs which are not homologous to the *am* genomic clone are removed from the membrane using high stringency washes while the *am* mRNA remains bound by virtue of its homology to the genomic clone. The *am* mRNA can be recovered from the membrane by denaturation of the mRNA-genomic clone duplex and the recovered mRNA can then be used for cDNA synthesis. The conditions used for this procedure are described in section 2.10 of the Materials and Methods. 40 µg of *am* genomic clone was immobilised on two small circles of Genescreen plus membrane (20 µg per circle) and used to isolate *am* mRNA from 75 µg of *N.crassa* mRNA (mRNA preparation c in tables 3.1 and 3.2) in a hybridisation selection reaction.

Expected Yield of *am* mRNA after Hybridisation Selection

Although 40 µg of p8-3, the *am* genomic clone, was immobilised on the Genescreen plus membrane for use in the hybridisation selection procedure, only 1.8kB of the 5.5kB p8-3 plasmid encodes the *am* gene. Therefore, the amount of

FIGURE 3.2

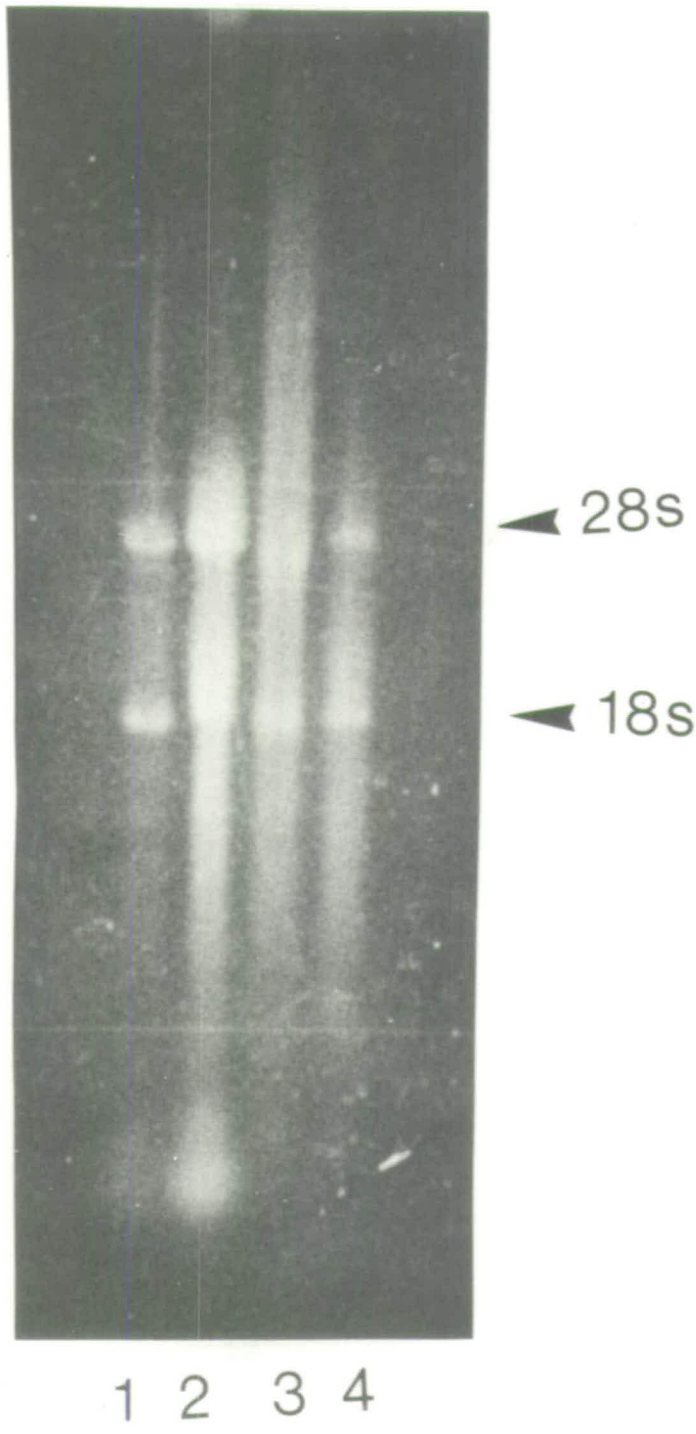


FIGURE 3.2

Example of *Neurospora crassa* total RNA and mRNA samples after electrophoresis on a 1.2 % agarose gel with formaldehyde. Samples are as follows:

Track 1: 5 µg *N.crassa* total RNA
Track 2: 10 µg *N.crassa am4* T-M RNA
Track 3: 12 µg *N.crassa* mRNA prep. C
Track 4: 10 µg *N.crassa* mRNA prep. B

mRNA preps. B and C refer to the same mRNA preparations detailed in tables 3.1 and 3.2.

am4 T-M RNA is total RNA which has been isolated from *N.crassa am4*, a strain of *N.crassa* which has its *am* gene deleted, and had its poly(A^+) RNA removed by oligo dT cellulose treatment.

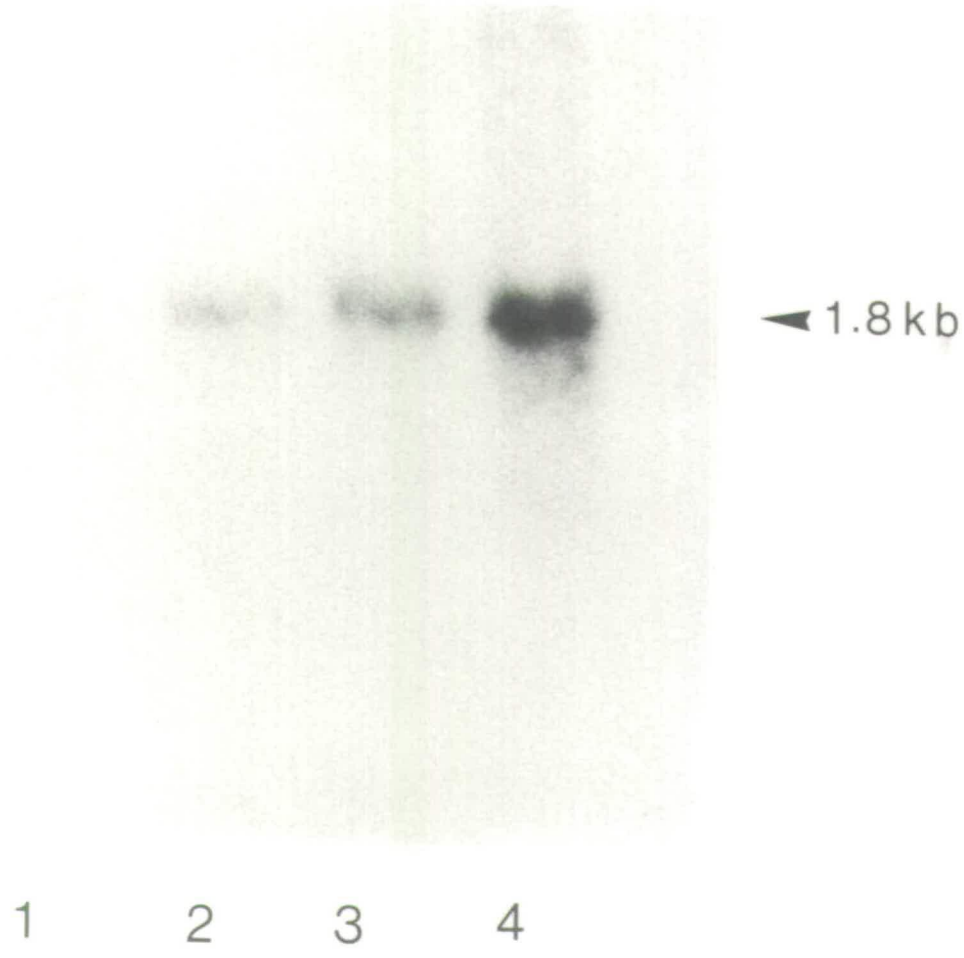
FIGURE 3.3

Overnight exposure of a Northern blot of a typical RNA gel, probed with the *am* genomic clone, p8-3, which contains the entire coding sequence of the *am* gene. Samples are as follows:

Track 1: 10 µg *N.crassa am4* T-M RNA
Track 2: 10 µg *N.crassa* total RNA
Track 3: 1 µg *N.crassa* mRNA prep. C
Track 3: 5 µg *N.crassa* mRNA prep. C

mRNA prep. C and *am4* T-M RNA are as described in the legend for figure 3.2.

FIGURE 3.3



DNA on the filters capable of hybridising to *am* mRNA was calculated as follows: Size of p8-3 = 5.5 kB
Size of DNA corresponding to *am* mRNA = 1.8 kB

$\therefore$ proportion of plasmid DNA corresponding to *am* mRNA = 1.8/5.5 = 0.33

Since the amount of DNA bound to the membrane = 40 µg, The amount of DNA corresponding to *am* mRNA = 40 X 0.33 µg = **13.3 µg.** 13.3 µg represents the amount of double stranded p8-3 DNA which corresponds to the area on the *am* genomic clone encoding the mRNA. However, only one strand of the DNA is capable of hybridising to the *am* mRNA, so the amount of DNA capable of hybridising to *am* mRNA will be 13.3/2 µg, i.e. 6.65 µg. Therefore, using the conditions described, a maximum of 6.65 µg of mRNA could be selected.

The *am* mRNA represents 0.1 % of *N.crassa* mRNA (Jane Kinnaird, personal communication) so the expected yield of *am* mRNA from 75 µg of *N.crassa* mRNA would be 75 ng. Since this figure is well below the maximum amount of *am* mRNA which could be selected using the hybridisation selection procedure, 100 % recovery of the *am* mRNA from the *N.crassa* mRNA should be possible.

An estimation of the actual yield of mRNA recovered from the hybridisation selection procedure was not made because the procedure was not expected to yield more than 75 ng, which was too small an amount to quantify accurately. For the purposes of further experiments involving this mRNA, the assumption was made that 100 % of the *am* mRNA was recovered. However, to ensure that the hybridisation selection procedure had been successful and that mRNA homologous to the *am* genomic clone had been recovered, a dot blot was carried out to check that the mRNA selected was capable of hybridising to ^{32}P labelled *am* genomic clone. The result of the dot blot is shown in figure 3.4. It can clearly be seen in figure 3.4 that a signal was evident in the dot corresponding to the hybrid selected material after only a 2 hour exposure. Signals in the other dots corresponding to non-selected mRNA and total RNA were not evident until an overnight exposure was carried out. This confirmed that the hybridisation

selected mRNA had bound specifically to the *am* genomic clone DNA on the filter discs and was not the product of non-specific binding to the Genescreen plus membrane.

One other point of interest which resulted from the dot blot shown in figure 3.4 was that the *am* genomic clone, p8-3, hybridised, not only to the expected samples on the dot blot, but also to the *am4* T-M RNA which should have represented a negative control. *am4* T-M RNA is total RNA from a strain of *N.crassa* which carries a deletion for the *am* gene and which has had its mRNA removed by oligo dT fractionation. As such, it was not expected to show homology to *am* DNA. Therefore, hybridisation of the *am* genomic clone to this RNA in a dot blot was somewhat unexpected. Further investigation of this showed that hybridisation of the *am* genomic clone consistently produced a signal in *am*4 T-M RNA in dot blots, although no signal was evident in Northern blots (see figure 3.3). A possible explanation for this may be that dot blots are much more sensitive than Northern blots since the RNA is concentrated in a single spot. It may be that several RNA species which hybridise weakly to p8-3, but are not seen on Northern blots, are seen on dot blots because they are concentrated on the same area on the filter.

The unexpected hybridisation of the p8-3 to *am*4 T-M RNA indicated that the mRNA selected by the hybridisation selection procedure was unlikely to be a homogeneous population of *am* mRNA molecules, especially since the conditions used during the hybridisation selection were not as stringent as those used for the Northern and dot blots. It was likely that the selected mRNA consisted of a selection of RNA species which were capable of hybridising to either the *am* or the pUC 8 section of the p8-3 plasmid. It would have been useful at this point to have checked some of the *am* selected mRNA on a Northern blot to check that the *am* mRNA was the predominant RNA species present but it was felt that since the amount of selected mRNA was limited, it would serve a more useful purpose as a template for cDNA synthesis. The remainder of the *am* selected mRNA was therefore used to produce the cDNA library described in section 3.2.5.

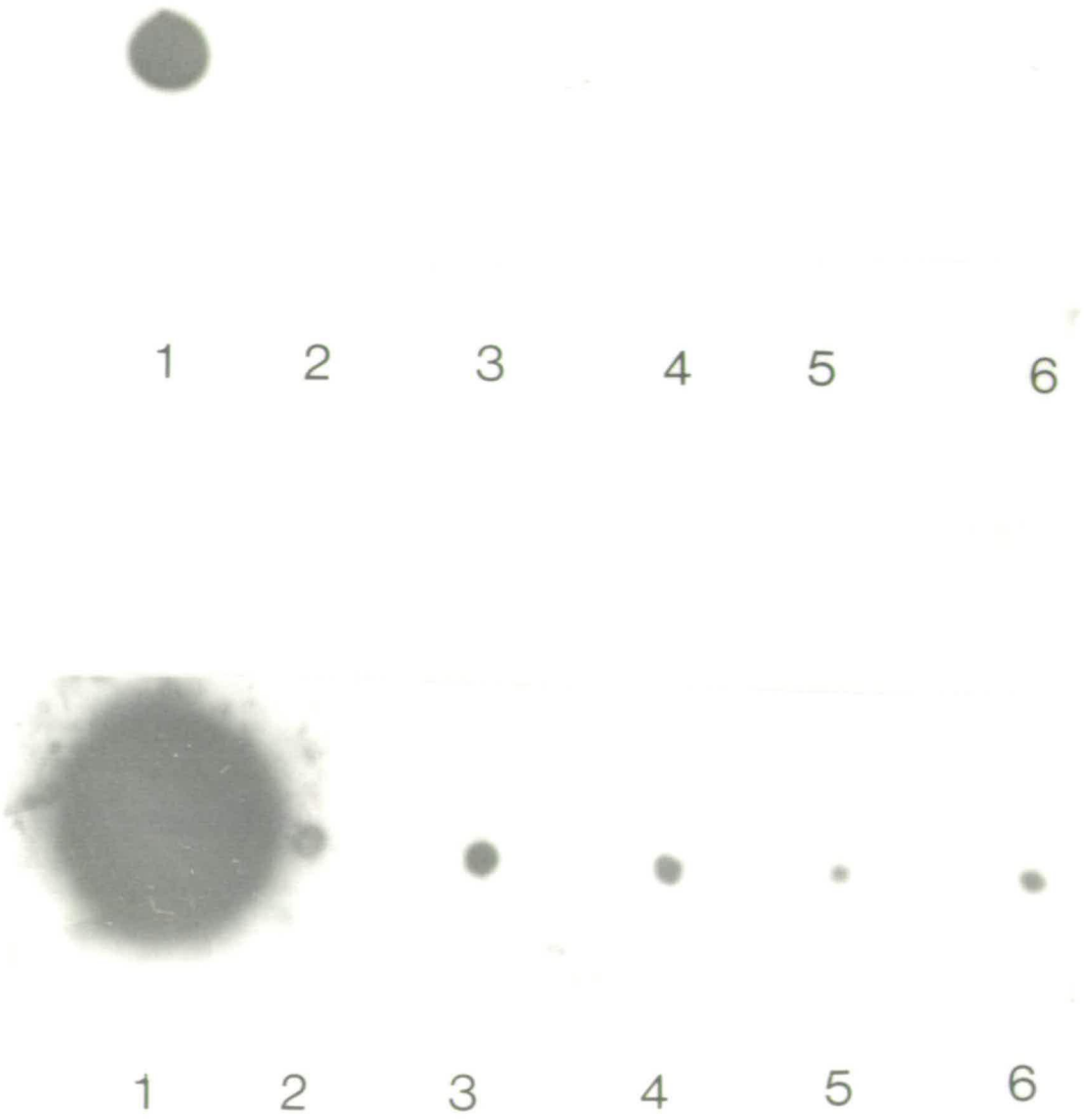
FIGURE 3.4

Autoradiographs of a dot blot, probed with the *am* genomic clone, p8-3, which contains the coding sequence of the *am* gene, following a two hour (top autoradiograph) and an overnight (bottom autoradiograph) exposure. Each sample was spotted onto the filter in a final volume of 4 µl. Samples are as follows:

1. 1 µl of *N.crassa am* mRNA isolated by hybridisation selection
2. 2 µg of *N.crassa* mRNA prep. A
3. 2 µg of *N.crassa* mRNA prep. B
4. 3 µg of *N.crassa* mRNA prep. C
5. 6 µg of *N.crassa* total mRNA
6. 10 µg of *N.crassa am-1* T-M RNA

mRNA preps A, B and C are as described in tables 3.1 and 3.2. *am-1* T-M RNA is as described in the legend to figure 3.2.

FIGURE 3.4



3.2.4 Optimisation of Conditions for cDNA Library Preparation

Conditions for cDNA Synthesis and Ligation of cDNA into Plasmid

As shown in table 3.2, a number of cDNA synthesis reactions were carried out to determine the optimum conditions for cDNA synthesis from total *N.crassa* mRNA before any attempts were made to synthesise cDNA from the *am*-selected *N.crassa* mRNA. This table shows clearly that the introduction of a 10 minute 70 °C incubation to denature any secondary structure present in the mRNA greatly improved the yield of first strand cDNA from mRNA (compare cDNA preparations 5 & 6 with 3 &4). It is also interesting to note that some consistency of 1st strand cDNA yield is seen with 2 of the 3 starting mRNA preparations, with preparation A producing 1st strand yields of 9.5 and 9.9 %, and mRNA preparation C showing yields of around 25-30 % for 3 out of the 4 cDNA synthesis reactions. It is not possible to draw any conclusions about consistency of 1st strand cDNA from mRNA preparation B because cDNA preparations 5 and 6 were made with the inclusion of a 70 °C incubation step whereas preparations 3 and 4 were made without a 70 °C step. mRNA preparation C consistently gave the best 2nd strand yields and produced cDNA of size range 200-5000 base pairs or greater (see figure 3.5). Attempts at size selection of the cDNA on agarose gels to enrich for the cDNA to the *am* gene were unsuccessful, producing very low numbers of recombinants (see table 3.2), and a successful cDNA library was only produced after ligating cDNA directly into vector immediately after synthesis and transforming the cDNA library into NM522 competent cells immediately after ligation. The low transformation efficiencies experienced with cDNA from synthesis reactions 7 and 8 were traced to inefficient ligations because the transformation efficiencies achieved were very similar to background levels achieved by transforming phosphatased *Sma* I cut vector, in the absence of insert. The observed transformants probably represent vector which has not been digested by *Sma* I. The high percentage of non-recombinants in cDNA preparations 6,9, and 10, all of which showed higher than background transformation efficiencies was traced to inefficient phosphatasing. This problem was overcome prior to cDNA preparation 2 which produced the cDNA library used in subsequent experiments.

In order to maximise the efficiency of the ligation of cDNA to phosphatased *Sma*

FIGURE 3.5

1 % agarose gel showing the size range of cDNA fragments produced after cDNA synthesis using 4.5 µg of *N.crassa* mRNA preparation 3. A predominant band of approximate size, 2.4kB, indicates the presence of an abundant mRNA species in the *N.crassa* mRNA. The samples are as follows:

Track 1: 1 µg of 1kB ladder (Gibco BRL)
Track 2: cDNA prepared in cDNA synthesis reaction 9 (see table 3.2)
Track 3: 1 µg of 1kB ladder (Gibco BRL)

The cDNA ranges in size from approximately 200 base pairs to greater than 4.5kB.

The size of the markers (in base pairs) in lanes 1 and 3 are as follows:

12216; 11198; 10180; 9162; 8144; 7126; 6108; 5090; 4072; 3054; 2036; 1635; 1018; 516/506; 396; 344; 298; 220; 201.

Three other marker bands of sizes 154, 134 and 75 base pairs are not evident.

FIGURE 3.5

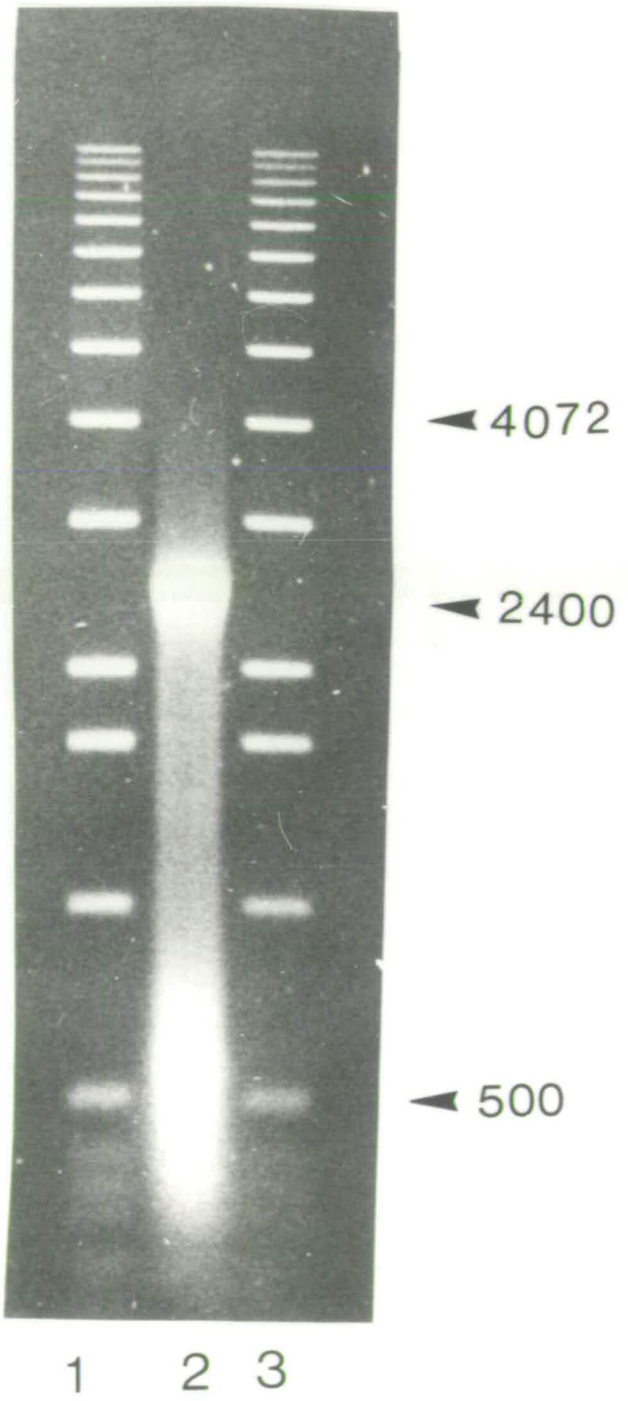


Table 3.2

Yields of 1st and 2nd Strand cDNA from *N.crassa* mRNA and Transformation Efficiencies.

mRNA preparation (concentration)	cDNA synthesis reaction	Yield of 1st strand cDNA	Initial 70°C incubation prior to 1st strand synthesis	Yield of 2nd strand cDNA	Size selection (size)	Transformation efficiency	%age non-recombinants	vector:insert ratio used in ligation
A (0.5 μg/μl)	1	9.9 %	YES	64.0 %	YES (1.0-4.0 kB)	ND	-	-
	2	9.5 %	YES	83.0 %	NO	3.4 X 10^5	5.6 %	3:1
B (1 μg/μl)	3	7.0 %	NO	19.0 %	NO	ND	-	-
	4	5.6 %	NO	ND	-	ND	-	-
	5	36.4 %	YES	0.0 %	-	ND	-	-
	6	13.3 %	YES	78.2 %	YES (1.5-3.5 KB)	3.6 X 10^5	99.3 %	1.5:1
C (1.5 μg/μl)	7‡	25.3 %	YES	104.0 %	YES (1.5-4.0 KB)	7.0 X 10^4	100.0 %	1:0.5
	8‡	25.2 %	YES	118.0 %				
	9	48.8 %	YES	102.0 %	YES (1.5-3.5 KB)	1.5 X 10^4	90.0 %	1:1.5
	10	29.0 %	YES	48.3 %	YES (1.3-3.5 KB)	1.6 X 10^5	98.3 %	1:1
						1.1 X 10^5	30.0 %	1:3.5

ND = not determined

* Transformation efficiency is given as colonies/μg of vector for all experiments except experiments 2 and 6, where the transformation efficiency is expressed as colonies/μg insert since the vector was in considerable excess in these experiments.

‡ The cDNA preparations from experiments 7 and 8 were pooled prior to size selection and ligation of the size selected cDNA into pTZ 18R.

Table 3.3

The Effect on Transformation Efficiency of Varying the Vector:Insert Ratio During Ligation

Source of cDNA	Vector:Insert Ratio (w:w)	Transformation efficiency (colonies/μg insert)	Non-recombinants (%)
Globin mRNA (control)	1:1	4.45×10^5	8.3 %
	1·5:1	3.17×10^5	10.6 %
	2·5:1	3.20×10^5	10.2 %
	3:1	7.04×10^5	11.8 %
	4:1	5.06×10^5	17.1 %
non-selected *N.crassa* mRNA	3:1	3.88×10^5	11.8 %
am-selected *N.crassa* mRNA	1:1	5.03×10^4	11.6 %
	1·5:1	2.73×10^5	11.7 %
	2:1	2.67×10^5	11.0 %
	3:1	4.30×10^5	16.3 %
	6:1	1.30×10^5	19.3 %

I cut vector, a number of different vector:insert ratios were used to determine the ratio which resulted in the highest transformation efficiency. The insert used to determine the optimum vector:insert ratio was a homogeneous population of cDNA synthesised from rabbit β-globin mRNA and ligations were carried out as described in section 2.12 of the Materials and Methods. The globin cDNA was prepared in the same manner as the *N.crassa* cDNA, with the exception of the 70 ^{o}C incubation prior to 1st strand synthesis, which was found to be unnecessary using globin mRNA, then immediately ligated into phosphatased *Sma* I cut pTZ18R. After overnight ligation at room temperature, the cDNA was transformed into competent NM522, prepared as described in section 2.6.1, using the Hanahan transformation procedure (see section 2.6.2). From Table 3.3, it is clear that a vector:insert ratio of 3:1 produced the highest transformation efficiencies using globin cDNA as insert. This ratio was therefore used as the vector:insert ratio for ligating cDNA prepared from non-selected *N.crassa* mRNA into phosphatased *Sma* I cut pTZ18R. Since it was possible that the quantity of cDNA prepared from *am*-selected mRNA may have been overestimated (see section 3.2.3), a number of vector:insert ratios were used when cloning this cDNA. The vector:insert ratios used are shown in table 3.3. Assuming that the calculation of cDNA yield for the *am*-selected mRNA was correct, the optimum ratio determined for *am*-selected cDNA proved to agree with the ratio of 3:1 determined for globin cDNA. The observed optimum vector:insert ratio of 3:1 for both globin and *am*-selected cDNA agreed with that suggested by King and Blakely, 1986, as the optimum vector:insert ratio for blunt end ligations using phosphatased blunt ended vector and phosphorylated insert.

Transformation of *E.coli* NM522 with pTZ18R

Initial transformation experiments using the $CaCl_2$ transformation procedure (see section 2.6.3) consistently produced transformation efficiencies of 5 X 10^5 transformants per µg uncut vector. Since transformation efficiency drops by a factor of 10-100 when *E.coli* is transformed with plasmid containing insert, $CaCl_2$ transformations would not produce high enough efficiencies for transforming cDNA libraries, so a protocol which would produce reproducible high efficiency transformation had to be found. In 1983, Douglas Hanahan had published a

comprehensive comparison of transformation methods and factors affecting transformation efficiency in *E.coli*. He suggested a protocol which was capable of producing transformation efficiencies of up to 10^9 transformants per µg plasmid with uncut plasmid. This procedure, although more complicated than $CaCl_2$ transformation, was therefore an ideal candidate for transforming cDNA libraries, contained in plasmid vectors, into *E.coli*. The original Hanahan procedure was optimised for several *E.coli* strains. NM522, however, was not one of them. Since the paper suggested that it may be necessary to alter some of the conditions slightly in order to optimise the procedure for other *E.coli* strains, some time was spent determining the conditions which yielded the best transformation efficiencies with NM522.

One of the conditions which was likely to require optimisation was the 42°C heat shock step. Hanahan originally suggested a 90 second heat shock. Therefore, in order to optimise the transformation efficiency which could be obtained when transforming pTZ18R into *E.coli* NM522, a number of transformation experiments were carried out using uncut pTZ18R and a time course from 0 to 120 seconds for the 42°C heat shock. The results from these experiments are shown in table 3.4. An additional observation made during the course of these experiments was that placing the cells on ice for 60 minutes prior to making them competent increased the transformation efficiency about 10 fold (compare experiments 2-7 with experiment 1 in table 3.4). Increasing the length of the ice incubation step to overnight (which can improve $CaCl_2$ transformations) resulted in a significant decrease in the transformation efficiency (compare experiment 8 with experiments 2-7 in table 3.4.).

From table 3.4., it can be seen that, for 5 of the 6 experiments in which the cells were incubated on ice for 60 minutes prior to being made competent, the optimum length of the heat pulse at 42°C was 60 seconds. Experiment number 4 showed results which did not agree with the pattern seen in experiments 2-3 and 5-7. The reason for the anomalous results in experiment 4 was not determined.

From the results in table 3.4., the transformation procedure described in sections 2.6.1 and 2.6.2 of the Materials and Methods was developed. Transformation efficiencies using this procedure consistently lay in the range of 10^7-10^8 with

Table 3.4

The Effect on Transformation Efficiency of Varying the Length of the 42°C Heat Pulse During Transformation of NM 522 with PTZ 18R.

Transformation Efficiency	Length of Heat Pulse (Seconds)									Length of Incubation of Cells on Ice Before Being Made Competent
	0	15	30	45	60	75	90	105	120	
1	–	1.50	1.06	1.10	1.50	1.65	1.90	1.00	1.36	0 minutes
2	2.80	9.10	8.90	7.45	12.00	5.30	7.30	7.67	5.08	60 minutes
3	–	13.00	12.00	12.40	13.60	8.40	7.70	5.50	3.80	60 minutes
4	2.69	8.00	14.00	12.00	8.40	9.30	11.00	7.70	6.85	60 minutes
5	20.00	23.00	24.60	25.70	29.00	20.40	18.70	21.60	23.20	60 minutes
6	23.00	29.80	30.20	32.40	37.00	20.40	23.20	28.00	27.40	60 minutes
7	0.31	0.65	1.50	1.40	1.70	1.10	1.60	1.80	1.80	overnight

Transformation efficiencies are expressed as number of colonies/μg uncut pTZ 18R. $\times 10^{-6}$.

uncut plasmid DNA, a range which, although lower than the 10^9 transformants per µg claimed by Hanahan, was considerably higher than the efficiencies achieved using the $CaCl_2$ transformation procedure.

3.2.5 Synthesis of cDNA from *am*-selected and Non-selected *N.crassa* mRNA

Having established the optimum conditions for cDNA synthesis, ligation of cDNA into vector and transformation of the cDNA library into NM522, cDNA from non-selected and *am*-selected mRNA were prepared. 3 µl samples of *am* -selected *N.crassa* mRNA and non-selected *N.crassa* mRNA (0.5 µg/µl) were used as templates for separate cDNA preparations. Following a 10 minute 70 ^{o}C incubation to dissociate any secondary structure in the mRNA prior to 1st strand synthesis, the cDNA synthesis was carried out according to the instructions provided with the cDNA synthesis kit used (see section 2.11). After synthesis, the cDNA was analysed for yield and quality of cDNA synthesised prior to ligation into pTZ18R and transformation into competent NM522 cells as described in section 3.2.6.

Analysis of the *am*-selected and non-selected cDNA after synthesis

To analyse the size and quality of the cDNA synthesised, 1st and 2nd strand cDNA samples from the reactions involving *am*-selected mRNA as template and non-selected mRNA as template were electrophoresed on a denaturing alkaline agarose minigel, prepared and run as described in sections 2.13.3 and 2.13.4. Since the cDNA is synthesised in the presence of α-^{32}P dCTP, the 1st and 2nd strand samples can be visualised by exposing the dried minigel to X-ray film. The resulting autoradiograph can be seen in figure 3.6. From this, it can be seen that the 2nd strand sample for each cDNA preparation showed the same size range as the 1st strand sample from the same preparation, indicating that no hairpin structures had been introduced during synthesis of the second strand cDNA. If hairpin structures had been present, the 2nd strand cDNA sample would produce a higher molecular weight band than the 1st strand sample on the alkaline agarose gel as the denaturing conditions provided by the gel cause the DNA to run as single stranded DNA rather than double stranded. It is important that there is no evidence of secondary structure in the cDNA molecules as this can create problems during the subsequent cloning procedures.

FIGURE 3.6

Autoradiograph of a denaturing alkaline gel showing 1st and 2nd strand cDNA samples reverse transcribed from *am*-selected and non-selected *N.crassa* mRNA, following overnight exposure.

The samples are as follows:

Track 1: Bacteriophage λ DNA digested with *Hin* dIII
Track 2: 1st strand cDNA reverse transcribed from non-selected mRNA
Track 3: 1st strand cDNA reverse transcribed from *am*-selected mRNA
Track 4: 2nd strand cDNA transcribed from non-selected 1st strand (track 2)
Track 5: 2nd strand cDNA transcribed from *am*-selected 1st strand (track 3)

The size of the λ/*Hin* dIII markers in track 1 are shown. Five other marker bands of sizes 23130, 2322, 2072, 564 and 125 are not evident in figure 3.6. The size of the band in tracks 3 and 4 was deduced from a longer exposure of the autoradiograph which did show the other marker bands.

FIGURE 3.6

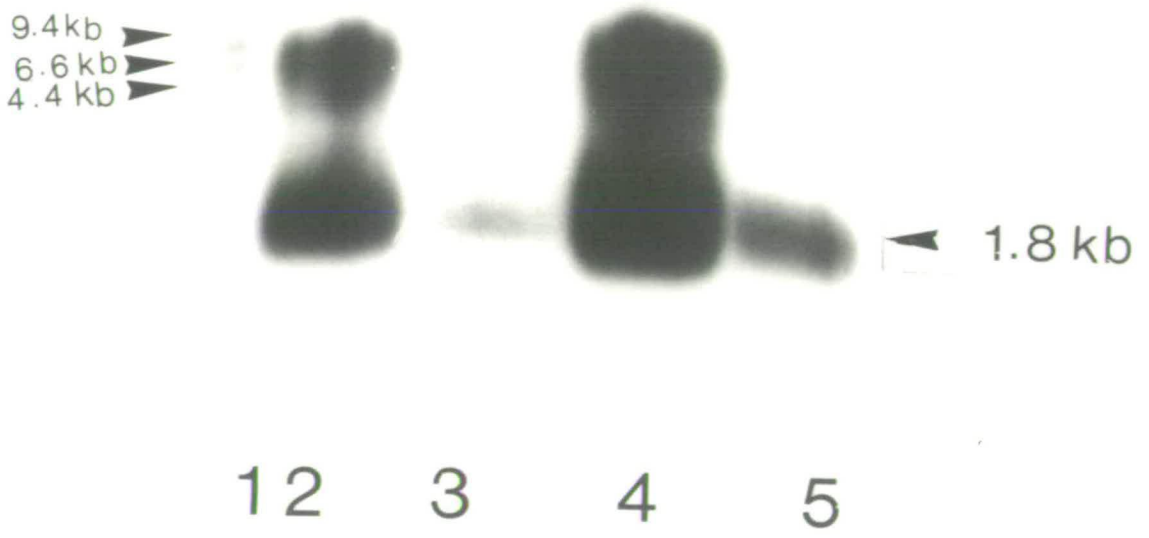


Figure 3.6 also shows that the 1st and 2nd strand cDNA derived from non-selected mRNA has produced a smear on the gel, indicating a mixture of cDNA species, whereas the 1st and 2nd strand cDNA derived from *am*-selected mRNA has produced 1 predominant band of approximate size 1.8kB, which compares very favourably to the size determined for the *am* mRNA on Northern blots (see figure 3.3).

Although alkaline agarose gel electrophoresis can be used to determine the size and quality of the cDNA, it does not provide any indication of the quantity of cDNA synthesised. This can, however, be determined from the amount of radioactivity incorporated into the first and second strand cDNA using the equation given in section 2.11.4 of the Materials and Methods. The percentage incorporation and estimated quantity of 1st and 2nd strand cDNA for non-selected and *am*-selected *N.crassa* mRNA are given in table 3.5.

Yield of cDNA from *am*-selected mRNA

Since the amount of *am*-selected mRNA used in the cDNA synthesis reaction had not been determined prior to cDNA synthesis, it was impossible to determine the yield of cDNA synthesised as a percentage of the starting mRNA. However, the amount of 1st strand cDNA synthesised from the *am*-selected mRNA could be estimated from the amount of radioactivity incorporated into nucleic acid during the first strand reaction. From table 3.5, the estimated amount of 1st strand cDNA synthesised from *am*-selected mRNA was 213 ng, and the yield of 2nd strand cDNA transcribed from 1st strand cDNA was 93 %. The amount of 1st strand cDNA synthesised (213 ng) was much higher than expected. However, since the amount of starting mRNA was unknown, it was likely that the amount of mRNA used in the cDNA synthesis reaction was higher than originally realised. This would indicate, therefore, that more mRNA had been selected during hybridisation selection than had been expected, either because the *am* mRNA represented greater than the expected 0.1 % of total *N.crassa* mRNA, or because the hybridisation selection procedure had selected other mRNA species which bore homology to the *am* genomic clone in addition to *am* mRNA. The latter explanation is more likely since the *am*-genomic clone, p8-3, used to select the mRNA, is capable of hybridising to *N.crassa* RNA which does not contain a copy of the *am* message (see section 3.2.3). However, the

Table 3.5

Yields of 1st and 2nd Strand cDNA from *am*-selected and non-selected mRNA and size of cDNA libraries produced.

	am-selected mRNA	non-selected mRNA
amount of mRNA used in reaction	ND	1.5 μg
% incorporation of radioactivity into 1st strand*	1.52 %	1.02 %
Quantity of 1st strand cDNA produced†	213 ng	143 ng
Yield of 1st strand cDNA reverse transcribed from mRNA†	ND	9.5 %
% incorporation of radioactivity into 2nd strand*	1.4 %	0.89 %
Quantity of 2nd strand cDNA synthesised†	198 ng	123 ng
Yield of 2nd strand cDNA transcribed from 1st strand cDNA†	93 %	86 %
Total no. of clones in 1° library	3.86×10^4	1.60×10^4

ND= not determined

* Incorporation of radioactivity into nucleic acid was calculated according to the method in section 2.11.4 of the Materials and Methods chapter.

† The quantity and yields of 1st and 2nd strand cDNA were calculated according to the equation given in section 2.11.4 of the Materials and Methods chapter.

presence of a single band in the *am*-selected 1st and 2nd strand cDNA lanes on the alkaline gel of similar size to that calculated for the *am* mRNA from Northern blots would seem to indicate that cDNA to the *am* mRNA was the predominant cDNA species.

Yield of cDNA from non-selected mRNA

Calculation of the yield of cDNA synthesised from non-selected *N.crassa* mRNA did not present any of the problems associated with the *am*-selected mRNA as the amount of starting material was known. Using the equation in section 2.11.4, it was estimated that 9.5 % of mRNA was reverse transcribed into 1st strand cDNA, yielding 143 ng of 1st strand cDNA. The amount of 2nd strand cDNA transcribed from 1st strand cDNA was 123 ng, a yield of 86 %.

The yield of 2nd strand cDNA in the non-selected *N.crassa* cDNA is lower than that achieved with cDNA prepared from *am*-selected mRNA. This may be related to the mRNA preparation used in the cDNA synthesis reaction. The mRNA used to prepare non-selected cDNA was mRNA preparation A, whereas the *am*-selected mRNA was derived from mRNA preparation C , a preparation which consistently produced better 1st and 2nd strand yields than either mRNA preparation A or mRNA preparation B (see table 3.2). The 1st strand yields from *am*-selected mRNA and non-selected mRNA cannot be compared because the yield of 1st strand cDNA from *am*-selected mRNA could not be determined.

3.2.6 Ligation of cDNA into pTZ18R and Transformation of cDNA Libraries into NM522

After synthesis of cDNA from *am*-selected and non-selected *N.crassa* mRNA, the cDNA was ligated into pTZ18R, which had been linearised with *Sma* I and phosphatased, using 1 unit of T4 DNA ligase (Gibco BRL) in the presence of 5 % PEG. The vector:insert ratios used were those described in table 3.3. After ligation overnight at room temperature, the ligation mix was transformed into competent NM522 cells according to the Hanahan procedure described in sections 2.6.1 and 2.6.2, which takes account of the modifications described in section 3.2.4. After transformation, half of each transformation reaction was plated on 137 mm agar plates and incubated overnight at 37 ^{o}C until colonies were formed. The second half of each transformation reaction was pelleted and

the cells resuspended in SOB medium containing 40 % glycerol for long term storage at -80 °C.

3.2.7 Storage and Screening of cDNA Libraries

After incubation of the agar plates containing the cDNA libraries, the plates were placed at 4 °C overnight. Permanent copies of the original plates were made on nitrocellulose discs as described in section 2.7 and these discs were used as masters. Replica filter discs were made from the masters as described in section 2.7 and these were used in colony hybridisation experiments, carried out as described in sections 2.15.4 and 2.16.3. The genomic clone of the *am* gene, p8-3, could not be used as a probe to find *am* cDNA clones within the library because the pTZ18R vector, into which the cDNA was cloned, contained considerable homology to pUC 8, the vector which contained the *am* genomic sequence. In order to generate a probe which would hybridise to *am* cDNA sequences and not to the vector DNA, the p8-3 plasmid was cut with *Cla* I and the 2038 base pair fragment which covers the coding region of the *am* gene was isolated by purification of the fragment after gel electrophoresis (see section 2.3). This 2kB *Cla* I fragment covers the entire coding region of the *am* gene and is completely free from pUC 8 sequence, so should not cross-hybridise onto the pTZ18R sequence. The 2 kB *Cla* I fragment was labelled with ^{32}P by nick translation, as described in section 2.14.1, and was used as a probe to identify *am* cDNA clones in colony hybridisation experiments (see section 2.15.4 and 2.16.3). Using the 2kB *Cla* I fragment as a probe, 19 potential *am* cDNA clones were identified in the non-selected *N.crassa* cDNA library and 1328 potential *am* cDNA clones were identified in the *am*-selected *N.crassa* cDNA library. Representation of *am* cDNA clones within the cDNA library had, therefore, been considerably enriched by using the hybridisation selected mRNA as a template for cDNA synthesis. The number of clones within the *am*-selected cDNA library capable of hybridising to the 2.0 kB *Cla* I fragment represent 9.95 % of recombinants in this library, whereas the number of clones within the non-selected cDNA library capable of hybridising to the 2.0 kB *Cla* I fragment represent only 0.11 % of recombinants. This indicates that hybridisation selection of *am* mRNA prior to cDNA synthesis has led to a 90-fold enrichment in the representation of *am* cDNA clones within a cDNA library.

The representation of *am* cDNA clones within the non-selected *N.crassa* cDNA library of 0.11 % agrees with the figure of 0.1 % suggested for the representation of *am* mRNA within *N.crassa* mRNA. This seems to infer that the cDNA library produced is a representative one. However, confirmation of this could only be achieved by probing the cDNA library for the presence of cDNA clones to an abundant mRNA species and comparing their representation within the library to that of their mRNA within *N.crassa* mRNA. Since this experiment was not done, no conclusions can be made about the representation of cDNA clones within the cDNA library.

The representation of *am*-positive cDNA clones within the *am*-selected library of 9.95 % confirms the theory proposed in section 3.2.4 that the unexpectedly high yield of 1st strand cDNA was due to the selection of RNA species, other than *am* mRNA, in the hybridisation selection procedure.

3.3 Summary

In this chapter, optimum conditions for cDNA synthesis and cloning were established. cDNA synthesis was improved by the introduction of a 70^{o}C incubation prior to 1st strand synthesis, which improved the yield of 1st strand cDNA transcribed from mRNA. The optimum conditions for cloning included ligation of the cDNA into vector immediately after synthesis; the inclusion of 5 % PEG in the ligation buffer; the use of a modified Hanahan transformation procedure to introduce the cDNA library into *E.coli* NM522 cells; and transformation of the cDNA library into the competent NM522 cells immediately after ligation. Modifications to the Hanahan transformation procedure included the introduction of a 60 minute incubation of the *E.coli* cells on ice prior to making the cells competent and reduction of the 42 ^{o}C heat pulse from 90 seconds to 60 seconds during the transformation procedure. Having established the best conditions for producing the library, mRNA, prepared from *N.crassa*, was used as a template for cDNA synthesis. Hybridisation selection of mRNA which was homologous to the *am* genomic clone, p8-3, prior to cDNA synthesis, led to a 90-fold enrichment of *am* cDNA clones within the cDNA library. A 2 kB fragment of the *am* genomic clone, p8-3, which covers the entire coding

region of the gene but does not include any plasmid sequence was isolated by gel electrophoresis following digestion with the restriction endonuclease, *Cla* I, and used as a probe to identify *am* cDNA clones. The total number of cDNA clones which produced a positive signal when the cDNA libraries were probed with this probe was 19 for the non-selected cDNA library and 1328 for the *am*-selected cDNA library. The isolation and characterisation of these clones will be dealt with in the next chapter.

Chapter 4 : Characterisation of am cDNA Clones

Chapter 4

Characterisation of *am* cDNA Clones

4.1 Introduction

In the last chapter, the synthesis and cloning of cDNA prepared from *N.crassa* mRNA samples which represented total mRNA from *N.crassa* and *N.crassa* mRNA which had been enriched for mRNA to the *am* gene by hybridisation selection was discussed. Screening the libraries using a 2.0 kB *Cla* I fragment of p8-3, which contains the coding sequence of the *am* gene, as a ^{32}P-labelled probe in colony hybridisation experiments identified 19 clones in the non-selected library and 1328 clones in the *am*-selected library which showed strong homology to this probe. This chapter will consider the purification and characterisation of these clones.

4.2 Results and Discussion

4.2.1 Isolation and Characterisation of Positive Clones from Non-selected *N.crassa* cDNA Library

Of the 3.88 X 10^4 cDNA clones in the primary non-selected *N.crassa* cDNA library, 19 showed a positive signal when probed with a 2kB *Cla* I fragment of the *am* genomic clone, p8-3, which covers the entire coding region of the *am* gene. Each of these clones was isolated and the recombinant plasmids purified by boiling lysis mini prep. as described in section 2.5.1. The size of the recombinant plasmid was determined by agarose gel electrophoresis, following restriction digestion with the restriction endonuclease *Hin* dIII. *Hin* dIII cuts pTZ18R only once and does not cut the *am* sequence. The size of the DNA band, after agarose gel electrophoresis, would therefore represent the sum of pTZ18R and the insert DNA. Since the size of the pTZ18R plasmid is known (2860 bases, Mead *et al.*, 1986), the size of the insert could be easily calculated.

Of the 19 cDNA clones isolated, none were full length. The size of the recombinant plasmids isolated from these clones proved to be surprisingly consistent and fell within the 3.3-3.4 kB size range, inferring an insert size of

around 400-500 base pairs. Confirmation of the plasmid size was obtained by digestion of the uncut DNA with *Bam* HI, an enzyme which cuts only once in pTZ18R and was not expected to cut in the cDNA sequence of the *am* gene. Electrophoresis of the *Bam* HI digested DNA produced bands which were the same size as those observed for the *Hin* dIII digests. Figure 4.1 shows the result of electrophoresis of *Hin* dIII and *Bam* HI digested recombinant plasmids from two of these positive clones.

Further restriction mapping of these 'short' *am* cDNA clones showed that the clones contained restriction sites common to the 3' end of the expected *am* transcript (figure 4.2). This confirmed that these clones were indeed *am* cDNA clones and not the result of cross-hybridisation of the *am* probe onto non-*am* cDNA clones.

There were four possible explanation for the consistent size of the short *am* clones isolated and the common restriction maps shown by the clones:

1. the clones isolated and characterised may have been derived from the same initial cDNA clone, leading to the incorrect assumption that all of the 19 positive clones were short;
2. the cDNA library had an insert size bias of 400-500 base pairs so only short cDNA molecules had been inserted into the plasmid;
3. the initial *am* cDNA had been full length but a large portion of the 5' end of the cDNA was commonly deleted on transformation of the recombinant plasmid into the *E.coli* host strain, NM522;
4. only short *am* cDNA molecules had been produced because the reverse transcriptase used to produce 1st strand cDNA had somehow been prevented from proceeding beyond the area corresponding to base pairs 1550-1650 of the *am* genomic sequence.

The first explanation was unlikely since each of the 19 clones was isolated from a cDNA library which had not been amplified, and, therefore, each clone would be represented only once in the library. Each of the 19 clones was treated separately throughout the isolation procedure and was cultured in such a way as to make selection of several copies of the same cDNA clone improbable.

Since it was unlikely that the short *am* cDNA clones were derived from the same original clone, the explanation for the short clones must lie in the construction of the cDNA library. Although it was unlikely that cDNA fragments of around 400bp in length had been preferentially cloned, this possibility was tested by

FIGURE 4.1

1 % agarose gel showing digestion of two *am* cDNA clones, isolated from the non-selected *N.crassa* cDNA library, with the restriction endonucleases, *Bam* HI and *Hin* dIII. The samples are as follows:

Track 1: 1 µg 1kB ladder (Gibco BRL)
Track 2: 1 µg pTZ18R digested with *Bam* HI
Track 3: 1 µg undigested *am* cDNA clone 1
Track 4: 1 µg *am* cDNA clone 1 digested with *Bam* HI
Track 5: 1 µg undigested *am* cDNA clone 1
Track 6: 1 µg *am* cDNA clone 1 digested with *Hin* dIII
Track 7: 1 µg undigested *am* cDNA clone 2
Track 8: 1 µg *am* cDNA clone 2 digested with *Bam* HI
Track 9: 1 µg undigested *am* cDNA clone 2
Track 10: 1 µg *am* cDNA clone 2 digested with *Hin* dIII
Track 11: 1 µg pTZ18R digested with *Hin* dIII
Track 12: 1 µg 1kB ladder (Gibco BRL)

The size of the single band seen in tracks 4, 6, 8 and 10 is approximately 3.3kB, as shown.

The size of the single band in tracks 2 and 11, corresponding to linear pTZ18R is 2.8kB, as expected.

The size of the marker bands in tracks 1 and 12 are as follows:

12216; 11198; 10180; 9162; 8144; 7126; 6108; 5090; 4072; 3054; 2036; 1635; 1018

FIGURE 4.1

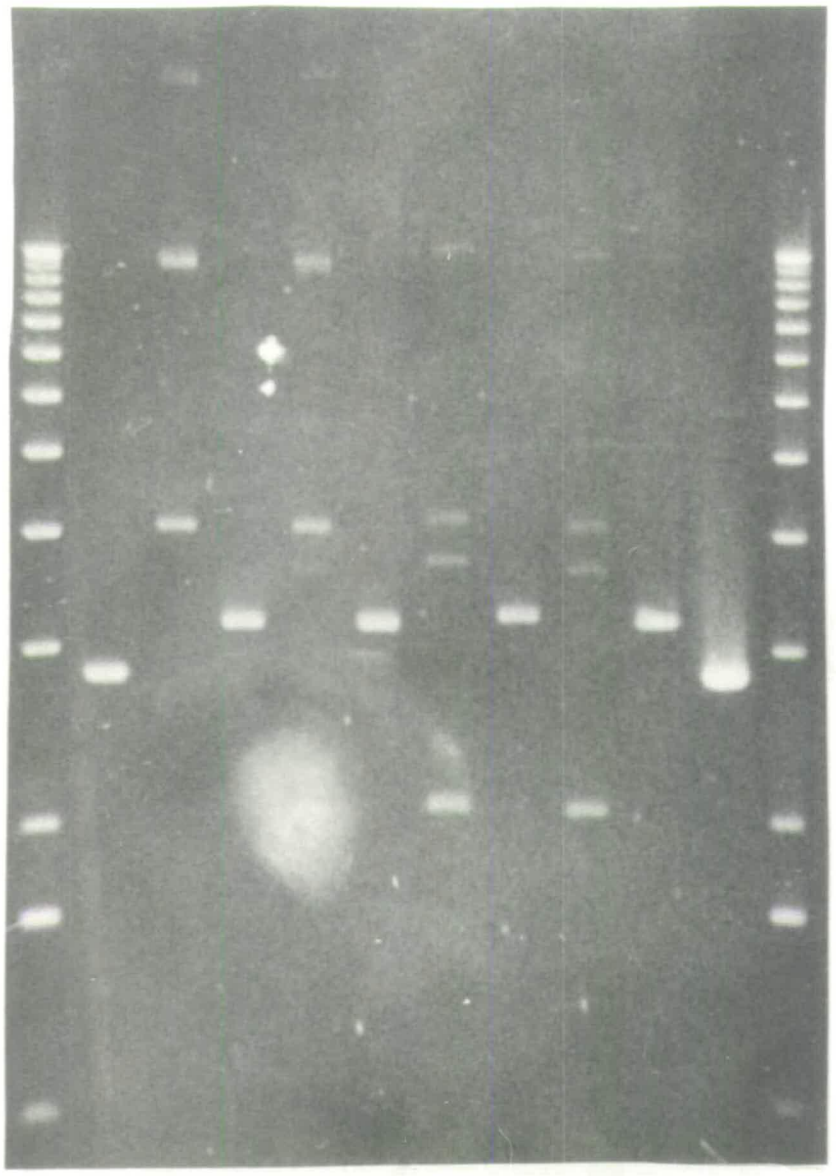
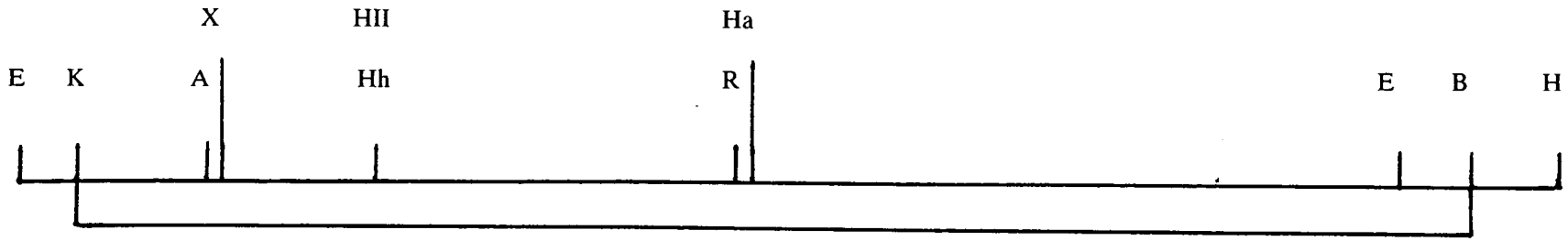


FIGURE 4.2



am insert DNA

= 100 base pairs

E=*Eco* RI; K=*Kpn* I; A=*Ava* I; X=*Xho* I; Hh=*Hha* I; HII=*Hae* II; R=*Eco* RII; Ha=*Hae* I; B=*Bam* HI; H=*Hin* dIII

isolating plasmid DNA from twenty recombinants which had been chosen at random from the non-selected *N.crassa* cDNA library, and determining the size of the recombinant plasmids by restriction analysis. The insert size range of these randomly chosen clones varied from 500 to 2000 base pairs, with no apparent bias towards smaller inserts. Size bias during the cloning procedure, therefore, was unlikely to be the explanation for short *am* cDNA clones.

To ensure that deletions were not generated in the recombinant plasmids during the colony isolation procedure, replica filters were made of the master filters containing the cDNA library, as described in section 2.7 and these were probed using a 235bp fragment of the *am* gene, generated by a double digest of the DNA with *Xho* I and *Bgl* II, which represents the area between the two introns at the 5' end of the coding sequence. Following hybridisation of the ^{32}P-labelled 235bp probe to the filters, no signal was evident upon autoradiography. This result confirmed that there were no full length *am* cDNA clones represented within the primary non-selected *N.crassa* cDNA library, and that, if deletion of the cDNA was the explanation for the presence of only short cDNA clones for the *am* gene in the cDNA library, then the deletion event must have taken place upon cloning the cDNA and was not a result of the colony isolation procedure. Deletion of DNA upon cloning into *E.coli* is normally sequence dependent (Greener, 1990, Fuchs *et al.*, 1988), so, if the explanation for the short *am* cDNA clones was deletion of the 5' end of the cDNA upon cloning, it was likely that this would also have been observed during the cloning of the genomic sequence. The *am* genomic clone, p8-3, did not show any DNA rearrangements after transformation into, and recovery of the DNA from, *E.coli* NM522, therefore deletion of the cDNA upon cloning was unlikely to be the explanation for the production of short *am* cDNA clones.

The most likely of the four explanations was that the reverse transcriptase was somehow prevented from proceeding beyond the region on the *am* mRNA, which corresponds to the 1550-1650bp area of the genomic sequence. The explanation for this did not lie in lack of processivity by reverse transcriptase since cDNA of 4kB and greater was synthesised during the reaction (see figures 3.5 and 3.6) so it could only lie within the mRNA structure itself.

Unfortunately time did not permit sequencing of the short *am* clones to determine exactly where the reverse transcriptase had stopped within the *am* sequence. However, since the restriction maps of the short clones indicated that the reverse transcriptase was not proceeding beyond the 1550-1650bp region, the structure responsible for preventing reverse transcription proceding beyond this point must lie 5' of this region. The most likely explanation for lack of processivity by reverse transcriptase was secondary structure within the mRNA. The observed improvement in 1st strand synthesis after the introduction of a 70 ^{o}C incubation step (see table 3.2) was consistent with this idea. It also indicated that *N.crassa* mRNA contains substantial secondary structure so it was possible that the *am* mRNA contained secondary structure . The high GC content of the *am* gene may make melting of hairpin loops and other secondary structure problematic, even at 70 ^{o}C. To identify any potential secondary structure at the 1550-1650bp region of the *am* mRNA, the sequence of the mRNA was put through UWGCG Fold and Squiggles programmes. The result of this is shown in figure 4.3. From this figure, it is clear that the potential to form extensive secondary structure exists within the region of the *am* mRNA sequence corresponding to the area 5' of the 1550-1650bp region. No such potential was found 3' of this area.

It would seem that the most likely explanation for the prevalence of short *am* cDNA clones within the non-selected *N.crassa* cDNA library was inefficient denaturing of secondary structure in the mRNA prior to cDNA synthesis.

4.2.2 Characterisation of *am* cDNA clones from the *am*-selected library

Having established that no full length *am* cDNA clones were contained within the non-selected library, it was important to ensure that the positive cDNA clones from the *am*-selected cDNA library were not also short cDNA clones. All of the colonies which produced positive signals when probed with the 2kB *Cla* I fragment also produced positive signals when probed with the 235bp *Xho/Bgl* fragment. This indicated that all of the potential positives in the *am*-selected cDNA library contained at least part of the 5' end of the mRNA transcript. It is possible that the smaller amount of mRNA used to make the *am*-selected cDNA library was more easily fully denatured by heating at 70 ^{o}C and so the potential

FIGURE 4.3

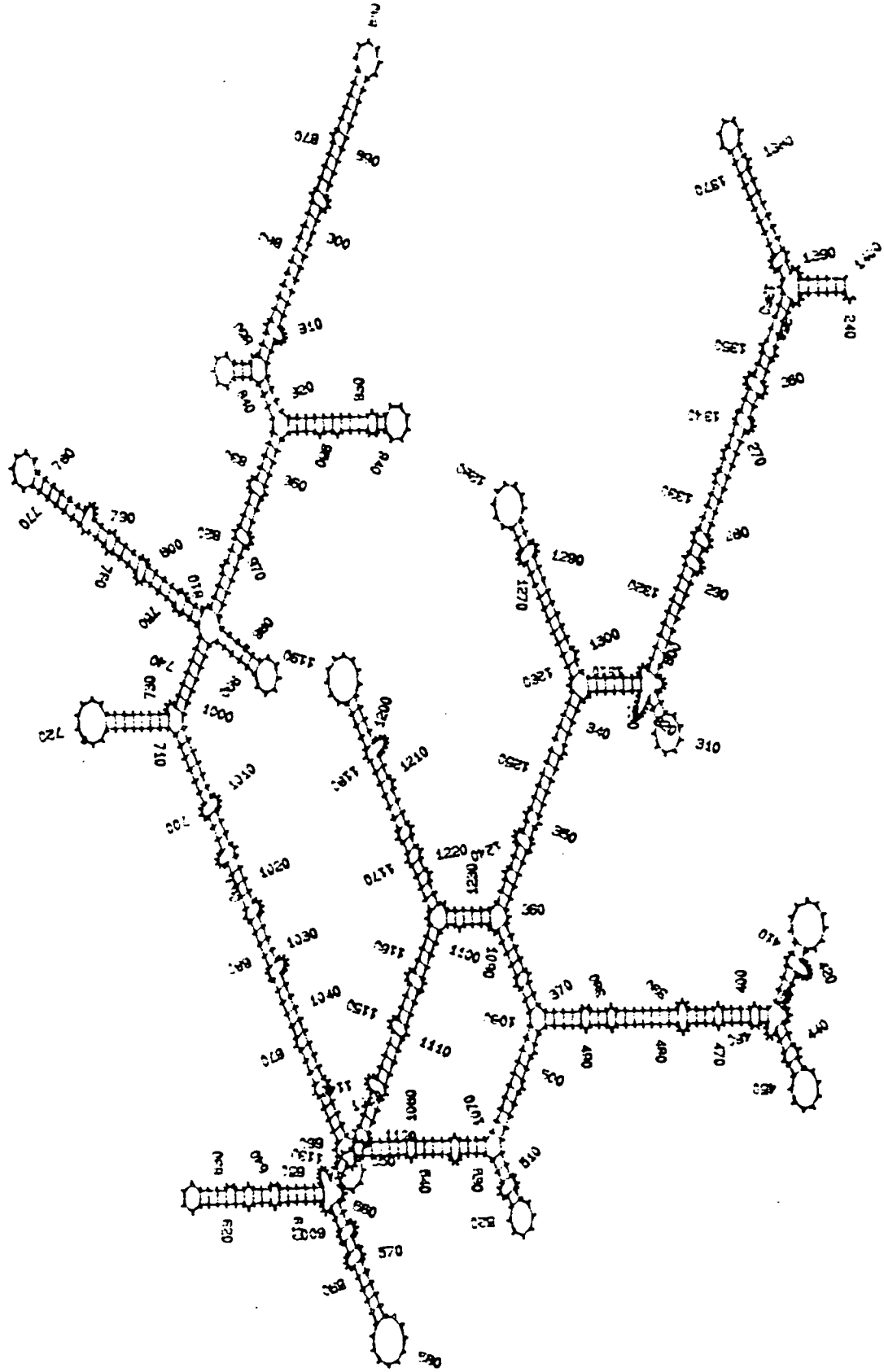


FIGURE 4.3

Potential secondary structure in the *am* mRNA generated using the University of Wisconsin Genetics Computer Group (UWGCG) sequence analysis packages 'FOLD' and 'SQUIGGLES'.

secondary structure problem evident in the non-selected library was avoided.

Having established that the positive *am* cDNA clones in the *am*-selected cDNA library appeared to be full length clones, a number of them were isolated and the size of the plasmid DNA determined by restriction digests. A number of the clones initially isolated produced an anomalous result when cut with the restriction endonuclease, *Bam* HI. As mentioned in the previous section, *Bam* HI was expected to cut only once within the recombinant plasmid, since it should cut only once in the pTZ18R sequence and should not cut at all in the *am* cDNA sequence, since both *Bam* HI sites in the *am* genomic sequence are outside the transcribed region of the gene. However, 5 of the *am* cDNA clones selected produced 2 bands, of sizes 2.8 kB and 1.6 kB when digested with *Bam* HI. Figure 4.4A shows *Bam* HI digests of 8 of the positive clones, including the 5 clones which produce 2 bands when digested with *Bam* HI. Although the digests shown in figure 4.4A are partial, bands at 2.8 and 1.6 kB can clearly be seen in lanes 6-10. With the exception of *Bam* HI, all other restriction enzymes used in size determination of the positive clones produced the banding pattern expected for an *am* cDNA clone in pTZ18R.

To confirm that the cDNA clones selected were indeed *am* cDNA clones, and attempt to locate the position of the second *Bam* HI site evident in lanes 6-10, the agarose gel shown in figure 4.4A was sandwich blotted as described in section 2.15.1. One of the filters produced was probed with ^{32}P-labelled 2kB *Cla* fragment of p8-3, which covers the entire coding region of the *am* gene, and the other was probed with ^{32}P-labelled pTZ18R. Figure 4.4B shows the autoradiograph of the filter which was probed with ^{32}P-labelled 2kB *Cla* fragment. From this figure, it can be seen that only the cDNA clones which cut twice with *Bam* HI produced a positive signal when probed with the 2kB *Cla* fragment of p8-3. Of the other clones digested, the DNA in track 4 had been extracted from a blue non-recombinant colony and was expected to produce a negative signal when probed with 2kB *Cla* fragment. The DNA in track 5, although derived from a white recombinant colony, produced a similar banding pattern to that in track 4, with a strong band at 2.8kB and no apparent smaller bands. The DNA in track 11 showed bands at 4.5kB, 6.4kB and 9.7kB. At least

FIGURE 4.4

Figure 4.4A
1 % agarose gel showing seven cDNA clones, which produced a signal when hybridised with the 2kB *Cla* I fragment of p8-3 (which covers the entire *am* coding sequence), and one clone which contained non-recombinant plasmid DNA, after isolation from the *am*-selected cDNA library and digestion with the restriction endonuclease *Bam* HI.

Figure 4.4B
Autoradiograph of a Southern blot of the agarose gel shown in figure 4.4A after hybridisation with the 2kB *Cla* I fragment of p8-3 and exposure overnight. Only the DNA in tracks 6-10 inclusive produced a signal with this probe. The 2.8kB band seen in these tracks in figure 4.4A does not hybridise, inferring that it is derived from vector DNA.

The samples in tracks 6-10 showed predominant bands at 1.6kB and 2.8kB in figure 4.4A. The larger bands in these tracks are the result of partial digestion.

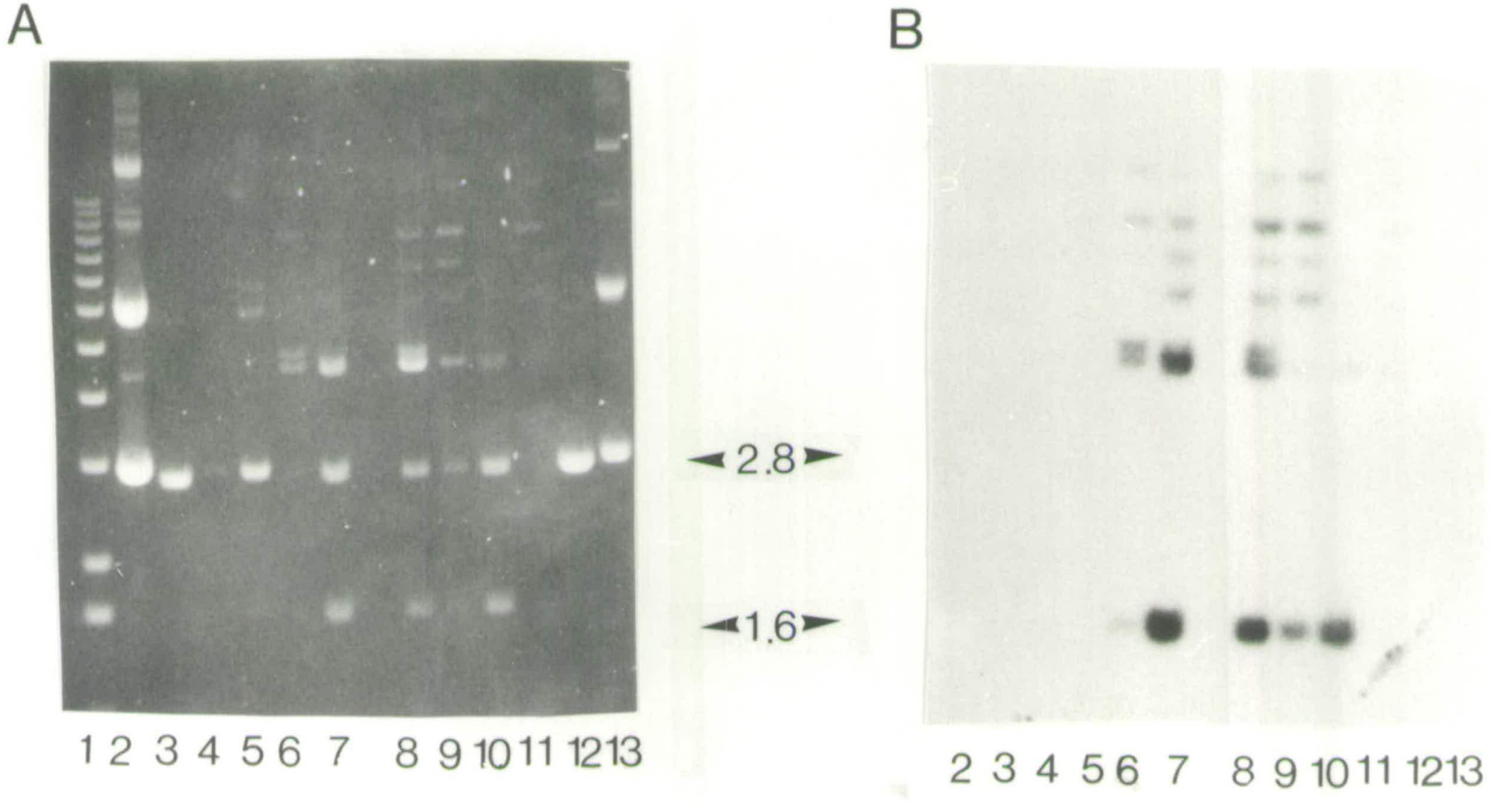
Samples are as follows, with approximate band sizes given in brackets for the tracks which did not produce a signal in figure 4.4B:

Track 1: 1 µg 1kB ladder (Gibco BRL)
Track 2: 1 µg undigested pTZ18R
Track 3: 1 µg pTZ18R digested with *Bam* HI (2.8kB)
Track 4: 100 ng non-recombinant DNA digested with *Bam* HI (5.9kB, 2.8kB)
Track 5: 1 µg cDNA clone 1 digested with *Bam* HI (6.8kB, 5.9kB, 2.8kB)
Track 6: 0.5 µg cDNA clone 2 digested with *Bam* HI
Track 7: 1 µg cDNA clone 3 digested with *Bam* HI
Track 8: 1 µg cDNA clone 4 digested with *Bam* HI
Track 9: 0.5 µg cDNA clone 5 digested with *Bam* HI
Track 10: 1 µg cDNA clone 6 digested with *Bam* HI
Track 11: 100 ng cDNA clone 7 digested with *Bam* HI (4.5kB, 6.4kB, 9.7kB)
Track 12: 1 µg pTZ18R digested with *Bam* HI (2.8kB)
Track 13: 1 µg undigested pTZ18R

The size of the marker bands in track 1 are as follows:

12216, 11198, 10180, 9162, 8144, 7126, 6108, 5090, 4072, 3054, 2036, 1635.

FIGURE 4.4



the larger two of the three bands are the result of partial digestion of the DNA, but the absence of a band at 2.8kB, the size of pTZ18R without insert, suggests that this cDNA clone contained an insert of approximate size 1.7kB, although that insert was not homologous to the *am* coding sequence. All of the bands seen in tracks 1-5 and 11-13 produced a signal when probed with ^{32}P-labelled pTZ18R (result not shown).

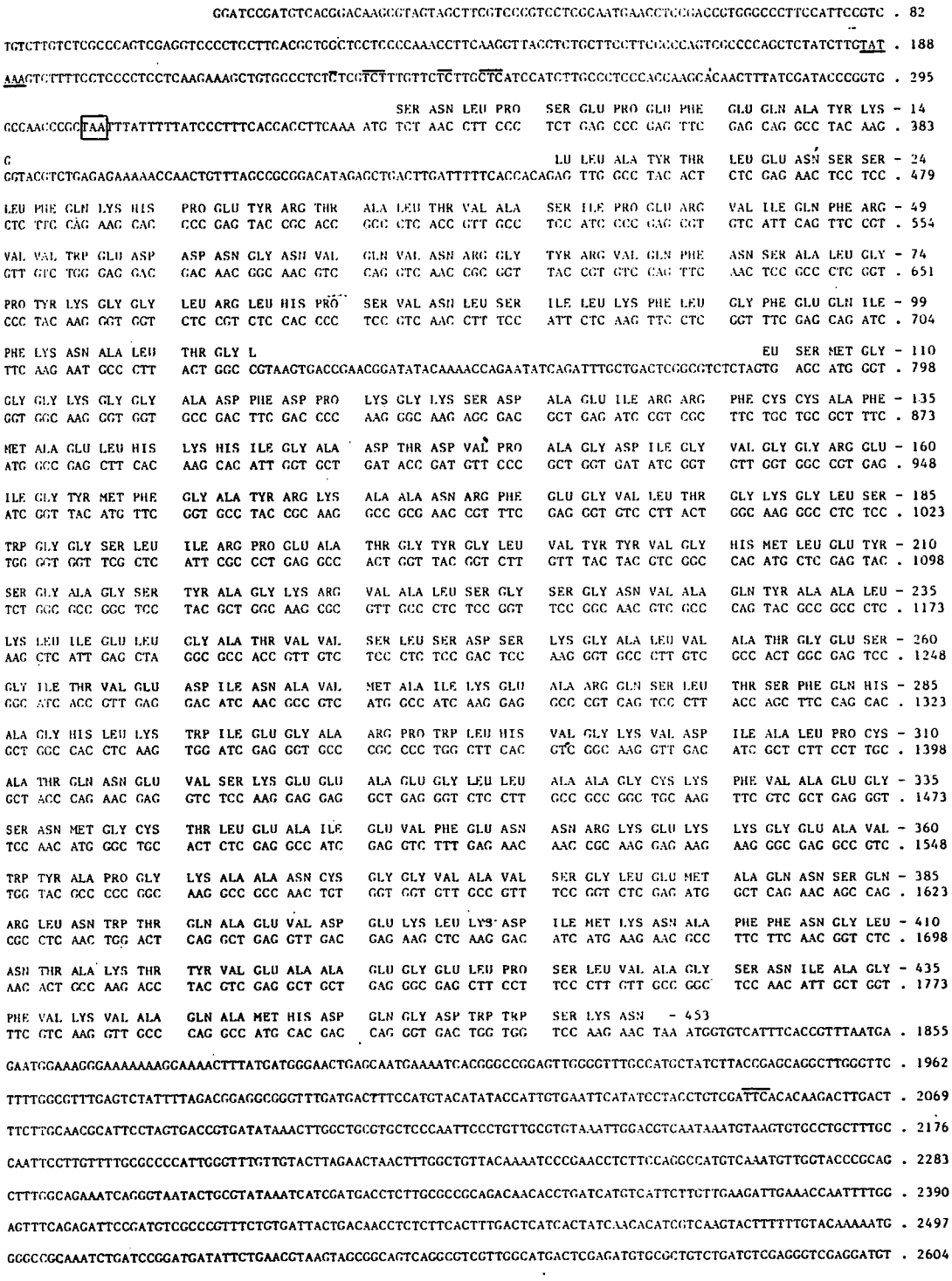
The DNA in tracks 6-10 did hybridise to the 2kB *Cla* fragment, with all of the bands which could be seen on the gel producing a signal, except the 2.8kB band. When probed with ^{32}P-labelled pTZ18R, all of the bands in tracks 6-10 which could be seen on the gel produced a signal, except the 1.6kB band (not shown). This confirms that the high molecular weight bands observed were the result of partial digestion of the recombinant plasmid DNA.

From the results of these hybridisations, it is clear that only the clones containing two *Bam* HI sites are true positives for *am*, since only the clones which produced two bands when cut with *Bam* HI produced a positive signal when probed with the 2kB *Cla* fragment of p8-3. These results also provide an indication to the position of the second *Bam* HI site within the cDNA clone. Since *Bam* HI appears to be capable of excising the insert from the plasmid, the second *Bam* HI site must be located at the 5' end of the insert as this is the end furthest from the *Bam* HI site on the polylinker of pTZ18R. The two most likely explanations for the occurrence of this unexpected *Bam* HI site were that a second small cDNA molecule containing a *Bam* HI site had been cloned into the plasmid as well as the *am* cDNA; or, that the *Bam* HI site had been created as a result of the cloning procedure. Since the cloning strategy had involved the blunt end ligation of the cDNA molecules into the plasmid vector, it was not impossible that more than one cDNA molecule could end up in the same vector. Indeed, the only precaution which was taken against this happening was to use an excess of vector to insert of 3:1. Since this was a weight:weight ratio and not a molar ratio, even this would not have guaranteed that there would only be one cDNA molecule in each plasmid molecule. The nature of blunt end ligations also opened up the possibility for the creation of new cloning sites. All four of the potential transcription initiation sites identified by Kinnaird and Fincham, 1983 (see figure

FIGURE 4.5

DNA sequence of the *Neurospora crassa am* gene, showing codons and deduced amino acid sequence for NADP-dependent glutamate dehydrogenase. The gene contains two introns which interrupt codons 15 and 107. The transcription start points and termination point identified by Kinnaird and Fincham, 1983 are overlined and the TATAAA sequence 5' of the start points is underlined. The TAA stop codon which is 5' of the initiating ATG is boxed.

FIGURE 4.5


GGATCCGATGTCACGGACAAGCGTAGTAGCTTCGTCCGGTCCTCGCAATGAACCTCCGACCGTGGCCCTTCCATTCCGTC . 82
TGTCTTGTCTCGCCCAGTCGAGGTCCCCTCCTTCACGCTGGCTCCTCCCCAAACCTTCAAGGTTAGCTCTGCTTCCTTCGCCCAGTGGCCCCAGCTCTATCTTGTAT . 188
AAAGTCTTTTCCTCCCCTCCTCAAGAAAGCTGTGCCCTCTCTCGTCTTTGTTCTCTTGCTCATCCATCTTGCCCTCCCACCAAGCACAACTTTATCGATACCCGGTC . 295
SER ASN LEU PRO SER GLU PRO GLU PHE GLU GLN ALA TYR LYS - 14
GCCAAGCCGCTAATTTATTTTTATCCCTTTCACCACCTTCAAA ATG TCT AAC CTT CCC TCT GAG CCC GAG TTC GAG CAG GCC TAC AAG . 383
LU LEU ALA TYR THR LEU GLU ASN SER SER - 24
GGTACGTCTGAGAGAAAAACCAACTGTTTAGCCGCGGACATAGAGCTGACTTGATTTTTCACCACAGAG TTG GCC TAC ACT CTC GAG AAC TCC TCC . 479
LEU PHE GLN LYS HIS PRO GLU TYR ARG THR ALA LEU THR VAL ALA SER ILE PRO GLU ARG VAL ILE GLN PHE ARG - 49
CTC TTC CAG AAG CAC CCC GAG TAC CGC ACC GCC CTC ACC GTT GCC TCC ATC CCC GAG CGT GTC ATT CAG TTC CGT . 554
VAL VAL TRP GLU ASP ASP ASN GLY ASN VAL GLN VAL ASN ARG GLY TYR ARG VAL GLN PHE ASN SER ALA LEU GLY - 74
GTT GTC TGG GAG GAC GAC AAC GGC AAC GTC CAG GTC AAC CGC GGT TAC CGT GTC CAG TTC AAC TCC GCC CTC GGT . 651
PRO TYR LYS GLY GLY LEU ARG LEU HIS PRO SER VAL ASN LEU SER ILE LEU LYS PHE LEU GLY PHE GLU GLN ILE - 99
CCC TAC AAG GGT GGT CTC CGT CTC CAC CCC TCC GTC AAC CTT TCC ATT CTC AAG TTC CTC GGT TTC GAG CAG ATC . 704
PHE LYS ASN ALA LEU THR GLY L EU SER MET GLY - 110
TTC AAG AAT GCC CTT ACT GGC CGTAAGTGACCGAACGGATATACAAAACCAGAATATCAGATTTGCTGACTCGGCGTCTCTAGTG AGC ATG GGT . 798
GLY GLY LYS GLY GLY ALA ASP PHE ASP PRO LYS GLY LYS SER ASP ALA GLU ILE ARG ARG PHE CYS CYS ALA PHE - 135
GGT GGC AAG GGT GGT GCC GAC TTC GAC CCC AAG GGC AAG AGC GAC GCT GAG ATC CGT CGC TTC TGC TGC GCT TTC . 873
MET ALA GLU LEU HIS LYS HIS ILE GLY ALA ASP THR ASP VAL PRO ALA GLY ASP ILE GLY VAL GLY GLY ARG GLU - 160
ATG GCC GAG CTT CAC AAG CAC ATT GGT GCT GAT ACC GAT GTT CCC GCT GGT GAT ATC GGT GTT GGT GGC CGT GAG . 948
ILE GLY TYR MET PHE GLY ALA TYR ARG LYS ALA ALA ASN ARG PHE GLU GLY VAL LEU THR GLY LYS GLY LEU SER - 185
ATC GGT TAC ATG TTC GGT GCC TAC CGC AAG GCC GCG AAC CGT TTC GAG GGT GTC CTT ACT GGC AAG GGC CTC TCC . 1023
TRP GLY GLY SER LEU ILE ARG PRO GLU ALA THR GLY TYR GLY LEU VAL TYR TYR VAL GLY HIS MET LEU GLU TYR - 210
TGG GGT GGT TCG CTC ATT CGC CCT GAG GCC ACT GGT TAC GGT CTT GTT TAC TAC GTC GGC CAC ATG CTC GAG TAC . 1098
SER GLY ALA GLY SER TYR ALA GLY LYS ARG VAL ALA LEU SER GLY SER GLY ASN VAL ALA GLN TYR ALA ALA LEU - 235
TCT GGC GCC GGC TCC TAC GCT GGC AAG CGC GTT GCC CTC TCC GGT TCC GGC AAC GTC GCC CAG TAC GCC GCC CTC . 1173
LYS LEU ILE GLU LEU GLY ALA THR VAL VAL SER LEU SER ASP SER LYS GLY ALA LEU VAL ALA THR GLY GLU SER - 260
AAG CTC ATT GAG CTA GGC GCC ACC GTT GTC TCC CTC TCC GAC TCC AAG GGT GCC CTT GTC GCC ACT GGC GAG TCC . 1248
GLY ILE THR VAL GLU ASP ILE ASN ALA VAL MET ALA ILE LYS GLU ALA ARG GLN SER LEU THR SER PHE GLN HIS - 285
GGC ATC ACC GTT GAG GAC ATC AAC GCC GTC ATG GCC ATC AAG GAG GCC CGT CAG TCC CTT ACC AGC TTC CAG CAC . 1323
ALA GLY HIS LEU LYS TRP ILE GLU GLY ALA ARG PRO TRP LEU HIS VAL GLY LYS VAL ASP ILE ALA LEU PRO CYS - 310
GCT GGC CAC CTC AAG TGG ATC GAG GGT GCC CGC CCC TGG CTT CAC GTC GGC AAG GTT GAC ATC GCT CTT CCT TGC . 1398
ALA THR GLN ASN GLU VAL SER LYS GLU GLU ALA GLU GLY LEU LEU ALA ALA GLY CYS LYS PHE VAL ALA GLU GLY - 335
GCT ACC CAG AAC GAG GTC TCC AAG GAG GAG GCT GAG GGT CTC CTT GCC GCC GGC TGC AAG TTC GTC GCT GAG GGT . 1473
SER ASN MET GLY CYS THR LEU GLU ALA ILE GLU VAL PHE GLU ASN ASN ARG LYS GLU LYS LYS GLY GLU ALA VAL - 360
TCC AAC ATG GGC TGC ACT CTC GAG GCC ATC GAG GTC TTT GAG AAC AAC CGC AAG GAG AAG AAG GGC GAG GCC GTC . 1548
TRP TYR ALA PRO GLY LYS ALA ALA ASN CYS GLY GLY VAL ALA VAL SER GLY LEU GLU MET ALA GLN ASN SER GLN - 385
TGG TAC GCC CCC GGC AAG GCC GCC AAC TGT GGT GGT GTT GCC GTT TCC GGT CTC GAG ATG GCT CAG AAC AGC CAC . 1623
ARG LEU ASN TRP THR GLN ALA GLU VAL ASP GLU LYS LEU LYS ASP ILE MET LYS ASN ALA PHE PHE ASN GLY LEU - 410
CGC CTC AAC TGG ACT CAG GCT GAG GTT GAC GAG AAG CTC AAG GAC ATC ATG AAG AAC GCC TTC TTC AAC GGT CTC . 1698
ASN THR ALA LYS THR TYR VAL GLU ALA ALA GLU GLY GLU LEU PRO SER LEU VAL ALA GLY SER ASN ILE ALA GLY - 435
AAC ACT GCC AAG ACC TAC GTC GAG GCT GCT GAG GGC GAG CTT CCT TCC CTT GTT GCC GGC TCC AAC ATT GCT GGT . 1773
PHE VAL LYS VAL ALA GLN ALA MET HIS ASP GLN GLY ASP TRP TRP SER LYS ASN - 453
TTC GTC AAG GTT GCC CAG GCC ATG CAC GAC CAG GGT GAC TGG TGG TCC AAG AAC TAA ATGGTGTCATTTCACCGTTTAATGA . 1855
GAATGGAAAGGGAAAAAAAGGAAAACTTTATGATGGGAACTGAGCAATGAAAATCACGGGCCGGAGTTGGGGTTTGCCATGCTATCTTACCGAGCAGGCTTGGGTTC . 1962
TTTTGGCGTTTGAGTCTATTTTAGACGGAGGCGGGTTTGATGACTTTCCATGTACATATACCATTGTGAATTCATATCCTAGCTGTCGATTCACACAAGACTTGACT . 2069
TTCTTGCAACGCATTCCTAGTGACCGTGATATAAACTTGGCTGCGTGCTCCCAATTCCCTGTTGCGTGTAAATTGGACGTCAATAAATGTAAGTGTGCCTGCTTTGC . 2176
CAATTCCTTGTTTTGCGCCCCATTGGGTTTGTTGTACTTAGAACTAACTTTGGCTGTTACAAAATCCCGAACCTCTTCCAGGCCATGTCAAATGTTGGTACCCGCAG . 2283
CTTTGGCAGAAATCAGGGTAATACTGCGTATAAATCATCGATGACCTCTTGCGCCGCAGACAACACCTGATCATGTCATTCTTGTTGAAGATTGAAACCAATTTTGG . 2390
AGTTTCAGAGATTCCGATGTCGCCCGTTTCTGTGATTACTGACAACCTCTCTTCACTTTGACTCATCACTATCAACACATCGTCAAGTACTTTTTTGTACAAAAATG . 2497
GGGCCGCAAATCTGATCCGGATGATATTCTGAACGTAAGTAGCCGCAGTCAGGCGTCGTTGGCATGACTCGAGATGTGCGCTGTCTGATGTCGAGGGTCGAGGATGT . 2604
TGCTGGATCACACTCGCGGGCACCGTACATACATGGATCC . 2644

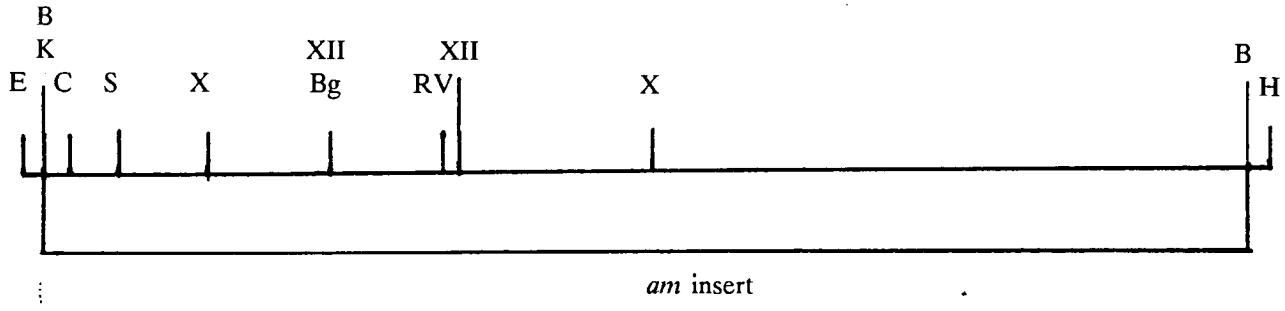
4.5), are 5' of the sequence ATCC, which has the potential to produce a *Bam* HI site (GGATCC) if ligated to a site with a 3' GG. Since *Sma* I, the enzyme used to produce the cloning site in pTZ18R for the cDNA libraries, leaves CCC as the 3' nucleotides, it did not seem likely that this was the explanation for the new *Bam* HI site. However, further restriction mapping of these clones revealed the apparent loss of the *Kpn* I site 5' of the *Sma* I site, which opened up a possibility for the creation of a *Bam* HI site without the intervention of a second cDNA molecule. If the *Kpn* I site had been lost as a result of the DNA ends being 'nibbled' during the alkaline phosphatase treatment (Maniatis *et al.*, 1982), it is not impossible that the 'nibbling' may have left a 3' GG, which, when ligated to the 5' ATCC of the *am* cDNA, would produce a new *Bam* HI site with the loss of only the *Kpn* I site from the polylinker (however, see section 4.2.4).

The occurence of several cDNA clones with 2 *Bam HI* sites was most likely an artefact of the isolation procedure used to purify the cDNA clones. Due to the large numbers of positive *am* cDNA clones, less precautions were taken to keep different positive clones on separate agar plates than had been taken with the clones from the non-selected *N.crassa* cDNA library so it was possible that some of the clones selected were derived from the same original cDNA clone.

Further restriction mapping of the clones containing two *Bam* HI sites revealed similar restriction patterns for all of them, which would suggest that they were all derived from the same original cDNA clone. A full restriction map of one of these clones, pAM-7, is shown in figure 4.6. From the map, it can clearly be seen that the cDNA insert in pAM-7 corresponds with the expected restriction map for an *am* cDNA clone which contains *am* in the correct orientation for expression. The 5' coding region of the *am* cDNA insert of pAM-7 was intact although part of the 3' end was missing, presumably because of incomplete second strand synthesis. However, since the important 5' coding region was present in the cDNA clone, it was decided to attempt to express the clone in the glutamate auxotroph, PA340.

4.2.3 Expression of pAM-7 in PA340

FIGURE 4.6



= 100 base pairs

E=*Eco* RI; K=*Kpn* I; B=*Bam* HI; C=*Cla* I; S=*Stu* I; X=*Xho* I; Bg=*Bgl* II; XII=*Xho* II; RV=*Eco* RV; H=*Hin* dIII

Maintenance of PA340

PA340 was maintained either on LB agar or on minimal M9N medium supplemented with the following: 0.1% thiamine; 40 µg/ml leucine; 40 µg/ml threonine; 40 µg/ml arginine; 40 µg/ml histidine; 2 mg/ml glutamate. The uptake of glutamate into *E.coli* cells is not very efficient, Berberich, 1972, therefore higher levels of glutamate must be added than that of the other amino acids, in order to maintain cell growth. In attempting to optimise the glutamate concentration to maintain the PA340 cells, I found that concentrations below 2 mg/ml glutamate produced very little growth on agar plates and no growth at all in liquid medium, whereas concentrations of glutamate above 2 mg/ml do not increase the level of cell growth above that produced by 2 mg/ml, and a decrease in cell growth is seen at glutamate levels above 50 mg/ml.

Transformation of pAM-7 into PA340

PA340 cells were made competent and the competent cells were transformed with 10 ng of pAM-7 DNA according to the method for $CaCl_2$ transformation of *E.coli* cells (see section 2.6.3). As a control, 10 ng of pTZ18R was also transformed into PA340 competent cells. Following transformation, 100 µl of the transformed cells were plated on LB agar plates supplemented with 75 µg/ml ampicillin and incubated overnight at 37 ^{o}C. Following overnight incubation, only one transformant was obtained for pAM-7 and 7 for pTZ18R, equivalent to transformation efficiencies of 1 X 10^3 and 7 X 10^3 colonies/µg respectively.

There were two reasons for the poor transformation efficiencies observed for PA340. Firstly, PA340 is known to be a poor transformation host and, secondly, PA340 does not carry any mutations in its DNA restriction and modification pathways. NM522, on the other hand, is *hsd*R$^-$ and *hsd*M$^-$, which means that DNA transformed into NM522 will not be recognised as foreign DNA and restricted. More importantly, DNA which has been isolated from NM522 will not have been modified in any way and so will still be vulnerable to restriction if placed in an *E.coli* strain which is not *hsd*R$^-$ (Lederberg, 1966, Sain and Murray, 1980). Since both pAM-7 and pTZ18R had been isolated from NM522 prior to transformation into PA340, it is likely that restriction of the unmodified plasmids accounts partly for the low transformation efficiencies observed. One way around this would have been to use an *E.coli* strain which was *hsd*R$^-$ but *hsd*M$^+$

either as the host strain for the library or as an intermediate strain to modify the DNA prior to transforming it into PA340.

Since pAM-7 was unmodified, it was also likely that the plasmid DNA may have been rearranged during transformation into PA340. To ensure that this was not the case, plasmid DNA was prepared from the pAM-7 clone in PA340 using a plasmid midiprep and the DNA was restriction mapped to ensure that the same restriction pattern was seen after pAM-7 had been in PA340 as had been evident prior to transforming the DNA into PA340. Having confirmed that the DNA had not undergone any changes, attempts were made to express pAM-7 in PA340.

Expression of pAM-7 in PA340

The PA340 cells containing pAM-7 were used to innoculate M9N liquid media supplemented with all possible combinations of glutamate and IPTG, i.e. glutamate alone; IPTG alone; glutamate plus IPTG; and neither glutamate or IPTG. PA340 cells containing pTZ18R were used as negative control in an identical set of experiments. Cell growth in the presence of IPTG alone would indicate that the *am* cDNA was in the correct frame for expression and that the gene product was capable of complementing glutamate auxotrophy in *E.coli*. If the cells only grow in the presence of glutamate, then it is likely that the *am* gene product is not being expressed. Table 4.1 shows that, for both sets of experiments, cell growth was only achieved in the presence of glutamate and neither pTZ18R or pAM-7 was capable of restoring glutamate prototrophy to PA340 in the absence of externally added glutamate, even when expression was induced by IPTG. Since the cells containing pAM-7 showed no growth restrictions in the presence of externally added glutamate, the only conclusion which could be drawn from this was that pAM-7 could not restore glutamate prototrophy to PA340. The most likely explanation for this was that the cDNA was inserted out of frame with the *Lac*Z gene.

Although it was impossible to determine the frame of the *am* cDNA with respect to the *Lac*Z reading frame without sequencing the DNA, the restriction map of pAM-7 did provide some useful information about the clone. The presence of the 5' *Cla* I site indicated that at least part of the 5' region which was transcribed

Table 4.1

Growth of PA340, transformed with pAM-7 and pTZ18R.

	PA340 transformed with pAM-7	PA340 transformed with pTZ18R
No IPTG or Glutamate	-	-
IPTG only	-	-
Glutamate only	+++	+++
IPTG plus glutamate	+++	+++

- = no growth +++ = growth

but not normally translated was present in the cDNA clone. Since the cDNA was being expressed as a fusion protein, this 5' region would be translated and would provide additional amino acids to the N-terminal portion of the NADP-GDH fusion protein, which may interfere with the activity of the protein in the *E.coli* cell. It was important, therefore, to determine whether this portion of the protein was creating any unforeseen problems with the expression of the *am* cDNA before any manipulation of the DNA to bring it into frame was done.

Further study of the sequence of this area of the *am* gene revealed an in frame TAA stop codon at a position 30 bases 5' of the ATG start codon of the *am* gene. Since this was likely to be a problem with any full length cDNA clone for the *am* gene, a way around it had to be found. The most straightforward way to change the sequence of the DNA would have been to mutate the DNA *in vitro* by using an oligonucleotide carrying a point mutation in the TAA stop codon to introduce a change in the DNA sequence at that position. However, this approach would not address any potential problem which might be caused by the additional amino acids introduced by the 5' sequence of the DNA which would not normally be translated *in vivo*. Since any manipulation of the DNA at this stage had to reduce the amount of extra sequence 5' of the coding sequence as well as bring the DNA into the correct reading frame and remove the TAA stop codon, it was decided to try to manipulate the cDNA sequence by enzymatic means to bring the DNA into frame.

After some initial unsuccessful attempts at controlled deletion of the 5' non-coding sequence using the exonuclease, Bal 31, it was decided that the best approach would be to look for restriction enzyme sites in the DNA which could be used to remove the 5' non-coding sequence and allow the DNA to be religated into the appropriate reading frame. At this stage a more detailed restriction map was necessary than the one shown in figure 4.6.

4.2.4 Detailed Restriction Map of pAM-7

Since the sequence of both pTZ18R and the *am* genomic clone were available, it was possible to identify potentially useful restriction sites for manipulating the *am* cDNA sequence and bringing the coding region into the same frame as the *LacZ* coding sequence. Since no unique restriction sites were available in the area

between the in frame TAA stop codon and the start of the *am* coding region, it became apparent that some loss of 5' coding sequence would be unavoidable using this procedure. Two restriction sites which should be unique within the *am* cDNA sequence and which were found at the 5' end of the *am* coding sequence were the restriction enzymes, *Ban* II, which cuts at position 360 of the *am* gene, and *Stu* I, which cuts at position 374 of the *am* gene sequence. *Ban* II looked to be the most useful of the two restriction sites since there was also a *Ban* II site in the polylinker 5' of the cloning site. Using the Mapsort programme of the UWGCG DNA sequence package to identify *Ban* II restriction sites in the pTZ18R sequence showed that the *Ban* II site in the polylinker at position 211 was unique within pTZ18R. Further investigation of these sites showed that it should be possible to bring the *am* coding sequence into frame with the *LacZ* coding sequence, if the cohesive termini left by digestion with *Ban* II were made blunt ended by filling in with the Klenow fragment of *E.coli* DNA polymerase I and religated together.

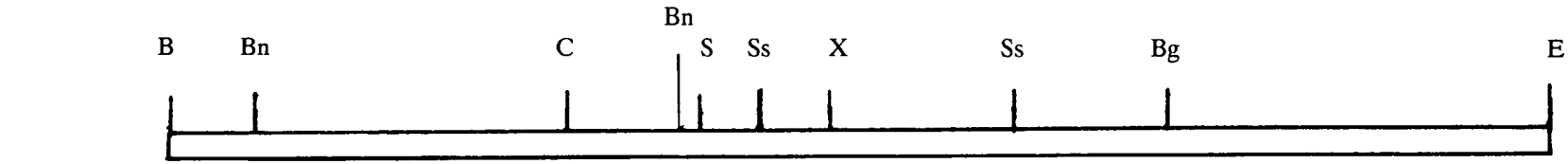
Digestion of pAM-7 with *Ban* II produced an unexpected result however, as three DNA fragments were produced of approximate sizes 2.3kB, 1.7kb and 300 base pairs. *Ban* II digestion of pTZ18R revealed the presence of 2 *Ban* II sites by producing 2 DNA fragments of approximate sizes, 2.3kB and 500 base pairs. Since the second *Ban* II site was not identified during the Mapsort and a published copy of the pTZ18R sequence was not available, it was necessary to use the original published sequences of pUC 19, Yanisch-Peron *et al.*,1985, and bacteriophage F1, Beck and Zink, 1981, from which pTZ18R was derived, and scan those by eye to identify the position of the second *Ban* II restriction site. Using this strategy, the second *Ban* II site was identified as being at position 2564 of the pTZ18R sequence. Since the 2.3kB band in pTZ18R and pAM-7 was derived from pTZ18R, the expected sizes of the DNA fragments generated from digestion within the insert of pAM-7 would be approximately 100 base pairs and 2kB. However, the DNA sizes of 1.7kB and 300 base pairs observed were significantly different from those expected. Since the orientation of the cDNA within the vector had already been established, the explanation for the DNA sizes observed did not lie in incorrect orientation of the insert. Further investigation involving electrophoresis of several µg of the *Ban* II digested pAM-7 on 1.5-1.8%

agarose gels revealed the presence of a fourth DNA fragment of approximately 140 base pairs in size. Since there is no *Ban* II site 3' of the site at position 360 of the *am* sequence, there had to be another *Ban* II site 5' of the known site, which suggested that the amount of DNA 5' of the *am* coding sequence was far greater than had been anticipated. The existence of a *Cla* I site in the expected position 3' of the coding sequence of *am* in pAM-7 suggested that the first strand cDNA synthesis had proceeded at least to that point. Since the most 5' transcription initiation signal identified by Kinnaird and Fincham, 1983, was located 58 bases 5' from the *Cla* I site, (see figure 4.5) the 5' end of the *am* cDNA must lie somewhere within that 58 bases. The additional DNA at the 5' end of the insert was therefore unlikely to be derived from *am*. To obtain more information about the 5' end of the insert DNA in pAM-7, it was necessary to map the DNA using restriction endonucleases.

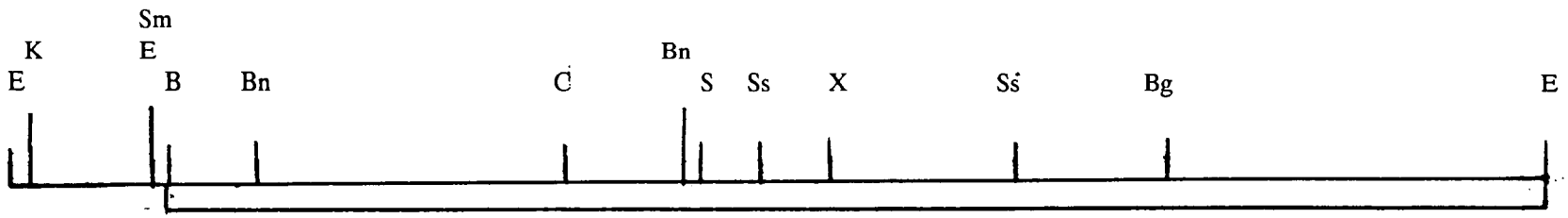
Although initial digests of pAM-7 with *Kpn* I had been unsuccessful (see section 4.2.2), manipulation of the digestion conditions proved that pAM-7 could indeed be digested with *Kpn* I. Since no problems had been encountered when digesting pTZ18R DNA with *Kpn* I, poor cutting of pAM-7 by *Kpn* I may be due to the close proximity of the *Kpn* I site to the cloning site. The position of a unique *Kpn* I site immediately 5' of the *Sma* I cloning site in pAM-7, allows the approximate size of the 5' non-coding region of the insert to be determined after digestion with *Kpn* I and a restriction enzyme which cuts at a site within the *am* sequence. Figure 4.8 shows the band sizes observed when pAM-7 is digested with *Kpn* I and *Bgl* II; *Kpn* I and *Cla* I; and *Bam* HI and *Cla* I. The map of pAM-7 shown in figure 4.7 shows the approximate size of the 5' non-coding region, calculated from the banding patterns shown in figure 4.8, as 400 base pairs. Comparison of the small DNA bands produced in tracks 4 and 6 of figure 4.8, which correspond to double digests with *Bam* HI and *Cla* I (track 4) and with *Kpn* I and *Cla* I (track 6) identified the position of the *Bam* HI site as approximately 80 base pairs 3' of the *Kpn* I site which was approximately 360 base pairs 5' of the *Cla* I site.

Double digestion of pAM-7 with *Ban* II and *Cla* I results in loss of the 300 base pair DNA fragment which suggests that the *Ban* II site in the 5' non-coding

FIGURE 4.7



am gene



vector

am insert from pAM-7

= 100 base pairs

E=*Eco* RI; K=*Kpn* I; Sm=*Sma* I; B=*Bam* HI; Bn=*Ban* II; C=*Cla* I; S=*Stu* I; Ss=*Sst* I; X=*Xho* I; Bg=*Bgl* II; RV=*Eco* RV

FIGURE 4.8

1.5 % agarose gel showing the restriction endonuclease digestions of PAM-7 DNA used to generate the map shown in figure 4.7.

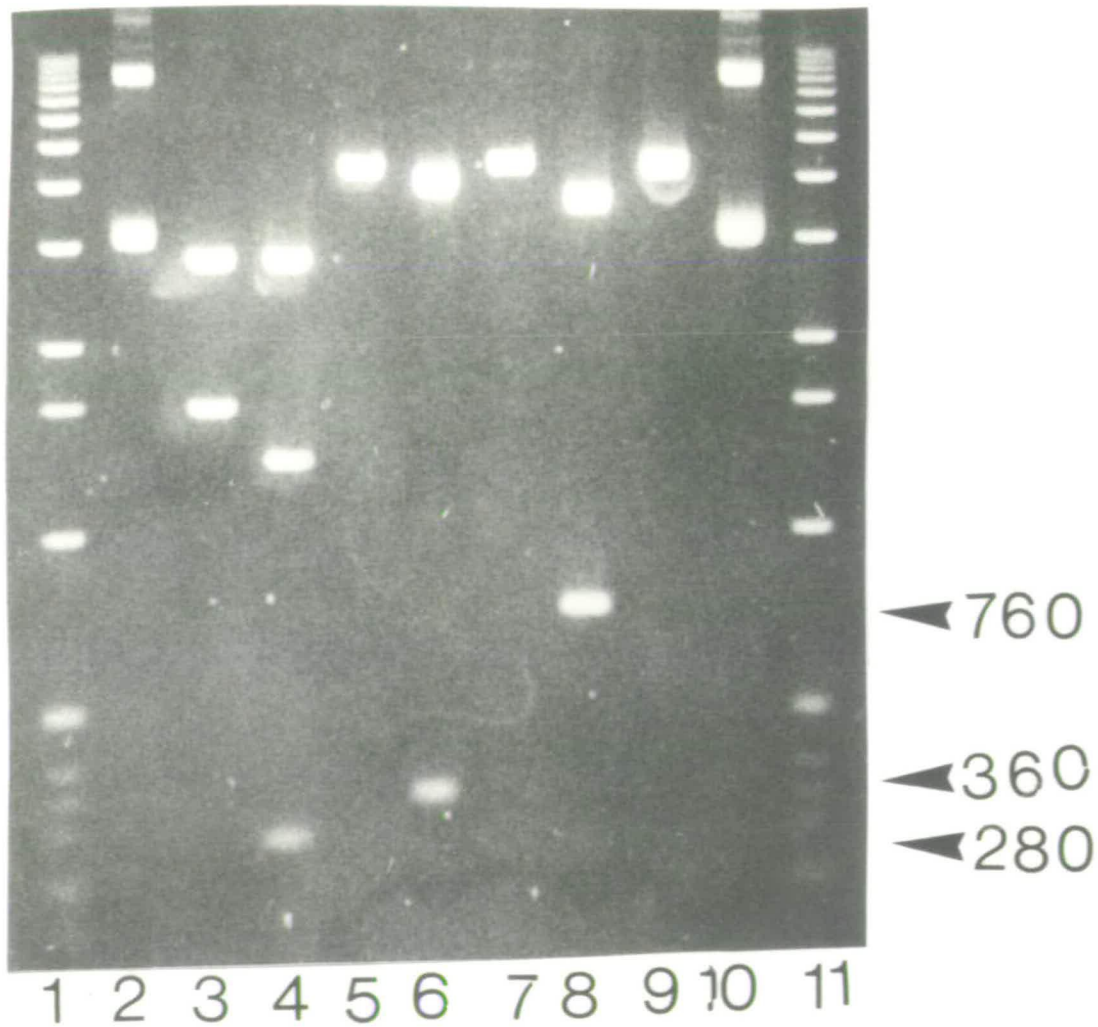
Unexpected band sizes which were generated are marked on the figure Samples are as follows with approximate band sizes given in brackets:

Track 1: 1 µg 1kB ladder (Gibco BRL)
Track 2: 1 µg undigested pAM-7
Track 3: 1 µg pAM-7 digested with *Bam* HI
(2.8kB, 1.6kB)
Track 4: 1 µg pAM-7 digested with *Bam* HI and *Cla* I
(2.8kB, 1.3kB, 280 bp)
Track 5: 1 µg pAM-7 digested with *Cla* I (4.4kB)
Track 6: 1 µg pAM-7 digested with *Cla* I and *Kpn* I
(3.9kB, 360bp)
Track 7: 1 µg pAM-7 digested with *Kpn* I (4.4kB)
Track 8: 1 µg pAM-7 digested with *Kpn* I and *Bgl* II
(3.5kB, 760bp)
Track 9: 1 µg pAM-7 digested with *Bgl* II (4.4kB)
Track 10: 1 µg undigested pAM-7
Track 11: 1 µg 1kB ladder (Gibco BRL)

The size of the marker bands in tracks 1 and 11 are as follows:

12216, 11098, 10181, 9162, 8144, 7126, 6108, 5090, 4072, 3054, 2036, 1635, 1018, 516/506, 396, 344, 298, 220, 201.

FIGURE 4.8



region of the insert is positioned 300 base pairs from the known *Ban* II site in the *am* coding sequence and 140 base pairs from the *Ban* II site in the polylinker of the plasmid. The position of a *Bam* HI site 280 base pairs 5' of the *Cla* I site and a *Ban* II site 220 base pairs from the *Cla* I site suggested another possible explanation for the 5' end of the cDNA clone which had not as yet been considered. Comparison of the deduced map of the 5' end of pAM-7 with the known genomic map of the *am* gene (see figure 4.7) indicates that the position of the *Bam* HI and *Ban* II sites in relation to the *Cla* I site on pAM-7 are almost identical to the position of those two sites in the *am* gene. In the *am* gene, there is a *Ban* II restriction site 68 base pairs 3' of the *Bam* HI site. Since the *Bam* HI site in pAM-7 was located approximately 80 base pairs 3' from the *Kpn* I site of the plasmid polylinker, the presence of the *Ban* II site from the *am* genomic sequence would explain the 140 base pair DNA fragment produced by *Ban* II digestion.

To confirm whether pAM-7 did indeed contain *am* genomic sequence, it was essential to determine whether or not the insert contained introns. Confirmation of the presence of introns in pAM-7 was obtained by double digestion of 3 µg of pAM-7 DNA with *Sst* II and *Eco* RV and separation of the DNA fragments on a high percentage agarose gel. Figure 4.9A shows the DNA fragments obtained after digestion of pAM-7 and p8-3 with *Sst* II and *Eco* RV. The presence of inserts in pAM-7 is confirmed by the appearance of DNA fragments of 331 and 178 base pairs in size after double digestion with *Eco* RV and *Sst* II. These fragments can be seen clearly in the tracks corresponding to *Eco* RV/*Sst* II digestions of both p8-3 and pAM-7. If pAM-7 had not contained the two introns of the *am* gene, then double digestion with *Eco* RV and *Sst* II would have produced only one small DNA fragment of size 270 base pairs and a large fragment of approximately 4.1kB. Figure 4.9B shows a more detailed comparison of p8-3 and pAM-7 following double digestion with *Bam* HI/*Eco* RV, *Bgl* II/*Stu* I, and *Eco* RV/*Stu* I. In each case, an identical band is produced for both p8-3 and pAM-7, which corresponds to the area containing the introns. The DNA fragment sizes which would have been produced if pAM-7 had not contained introns are given in the legend to figure 4.9B.

FIGURE 4.9

Figure 4.9A
1.8 % agarose gel indicating the presence of introns in the insert of pAM-7 by the production of restriction fragments of 331 and 178 base pairs in length after digestion with *Eco* RV and *Sst* II, which match those produced by the genomic clone, p8-3. The samples are as follows with approximate band sizes given in brackets:

1: 1 µg 1kB ladder (Gibco BRL)
2: 1 µg undigested p8-3
3: 2 µg p8-3 digested with *Eco* RV (5.3kB)
4: 2 µg p8-3 digested with *Eco* RV and *Sst* II
(4.9kB, 331bp, 178bp)
5: 2 µg pAM-7 digested with *Eco* RV and *Sst* II
(3.9kB, 331bp, 178bp)
6: 2 µg pAM-7 digested with *Eco* RV (4.4kB)
7: 1 µg undigested pAM-7
8: 1 µg 1kB ladder (Gibco BRL)

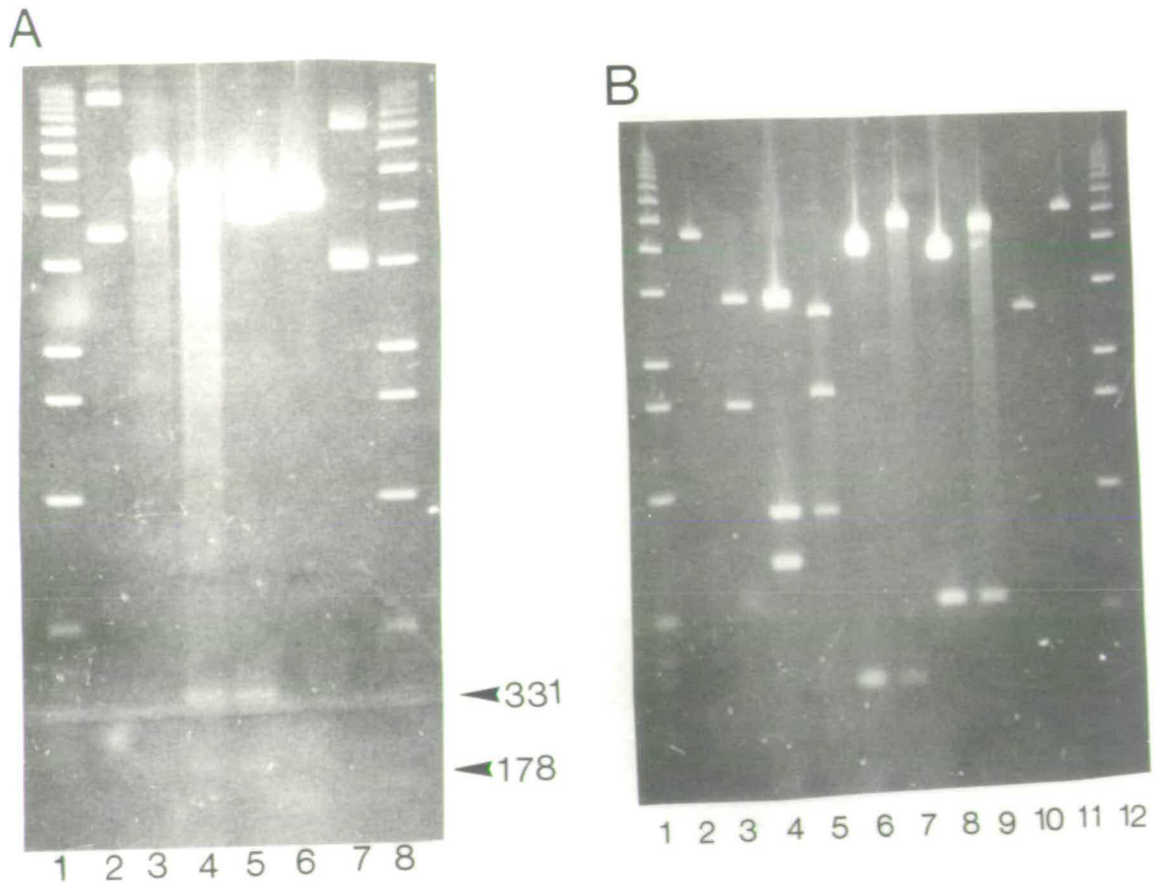
Figure 4.9B
1.5 % agarose gel showing detailed comparison of pAM-7 and p8-3 which confirms the presence of introns in pAM-7. Samples are as follows with approximate band sizes in brackets (expected sizes if introns were absent in pAM-7 are given in italics for tracks 4, 6 and 8):

1 : 1 µg 1kB ladder (Gibco BRL)
2 : 1 µg pAM-7 digested with *Bgl* II (4.4kB)
3 : 1 µg pAM-7 digested with *Bam* HI (2.8kB, 1.6kB)
4 : 1 µg pAM-7 digested with *Bam* HI and *Eco* RV
(2.8kB, 900bp *799*, 750bp)
5 : 1 µg p8-3 digested with *Bam* HI and *Eco* RV (2.7kB, 1.8kB, 900bp)
6 : 1 µg pAM-7 digested with *Bgl* II and *Stu* I (4kB, 340bp *276bp*)
7 : 1 µg p8-3 digested with *Bgl* II and *Stu* I (5kB, 340bp)
8 : 1 µg pAM-7 digested with *Eco* RV and *Stu* I
(3.8kB, 570bp *442bp*)
9 : 1 µg p8-3 digested with *Eco* RV and *Stu* I (4.7kB, 570bp)
10 : 1 µg p8-3 digested with *Bam* HI (2.7kB band represents two bands)
11 : 1 µg p8-3 digested with *Bam* HI (5.3kB)
12 : 1 µg 1kB ladder (Gibco BRL)

The sizes of the marker bands in tracks 1 and 8 of figure 4.9A and 1 and 12 of figure 4.9B are as follows:

12216, 11198, 10180, 9162, 8144, 7126, 6108, 5090, 4072, 3054, 2036, 1635, 1018, 516/506, 396, 344, 298, 220, 154, 134, 75.

FIGURE 4.9



Two possible explanations for the presence of introns in pAM-7 were considered:

1. the mRNA used to make the cDNA library had somehow been contaminated with genomic DNA
2. the insert of pAM-7 was derived from precursor mRNA which had not yet been spliced to form mature mRNA.

Although it is not uncommon to find cDNA reverse transcribed from precursor mRNA in cDNA libraries and it can represent up to 10% of the clones in a library (Andrew Hicks, MRC Molecular Neurobiology Unit, Cambridge, personal communication), the latter explanation seemed unlikely because of the large extent of *am* non-coding sequence which had been conserved in pAM-7 and which was situated 5' of the transcription initiation sites postulated by Kinnaird and Fincham, 1983. It is possible that the presence of another transcription initiation site positioned 5' of an intron in the non-coding region of the *am* gene could explain 5' non-coding sequence in a precursor mRNA for *am*, but this would require the presence of an intron splice site 3' of the transcription initiation sites postulated by Kinnaird and Fincham, 1983, and there was no evidence for an intron splice site between the postulated transcription initiation sites and the first intron of the *am* coding sequence.

Digestion of the original mRNA preparation with RNase H to identify any DNA present in the preparation showed no evidence of DNA contamination of the mRNA so, if pAM-7 was the result of genomic contamination of the mRNA preparation, the contamination must have taken place at a later stage. The most likely stage of the preparation at which this occurred was during hybridisation selection of mRNA for the *am* gene (see section 3.2.3). Since there was additional DNA 5' of the *Bam* HI site of the 5' non-coding region of *am* in pAM-7, this must be derived from the pUC-8 segment of the p8-3 plasmid, if the pAM-7 insert was the result of contamination of the hybridisation selected *am* mRNA by p8-3. Immediately 5' of the *Bam* HI site, containing the insert of p8-3, there is a *Sma* I site. Since the *Sma* I site of pTZ18R was destroyed by the introduction of insert DNA in pAM-7, digestion of pAM-7 with *Sma* I would suggest that this site was derived from p8-3. p8-3 also contains an *Eco* RI site 10 bases 5' of the *Bam* HI cloning site. Since the polylinker of pTZ18R contains an *Eco* RI site 18 base pairs 5' of the *Sma* I cloning site, digestion of pAM-7 with

Eco RI should produce two fragments of approximate sizes 4.3kB and 90 base pairs. Digestion of pAM-7 with these enzymes and double digests using *Eco* RI and *Cla* I and with *Sma* I and *Cla* I produced the results expected if the insert of pAM-7 was derived from p8-3, as shown in the following map:

-- -- ----E---------E-S-B----------C------- -- --

10 base pairs = -
E=*Eco* RI; S=*Sma* I; B=*Bam* HI; C=*Cla* I

The most likely explanation for the insert of pAM-7 was, therefore, that the insert was derived from part of p8-3, the *am* genomic clone, which had been introduced into the mRNA preparation during the hybridisation selection procedure. Since the presence of introns in pAM-7 meant that expression of the gene product in *E.coli* would be impossible and it could not be used for the expression studies, no further investigation of this clone was done.

4.2.5 Characterisation of Other *am* Positive Clones Isolated from the *am*-selected *N.crassa* cDNA Library

Having ruled out pAM-7 as a potential cDNA clone for the NADP-GDH expression studies, a further 148 cDNA clones which had produced a positive signal when probed with the 2kB *Cla* I fragment of p8-3, were isolated and their plasmid DNA analysed by restriction analysis. The DNA was digested with *Bam* HI and with a combination of *Stu* I and *Eco* RV to determine whether the two introns and 5' non-coding sequence of p8-3 were present. Clones which did not contain introns should produce a DNA fragment of 425 base pairs when digested with *Stu* I/*Eco* RV, rather than the fragment of 552 base pairs observed for the same digest of p8-3 and pAM-7. Clones which produced two fragments with *Bam* HI and/or a 552 base pair fragment with *Stu* I/*Eco* RV must be derived from genomic *am* DNA and so would not be useful for the expression studies. All of the 148 clones isolated produced a 552 base pair fragment when digested with *Stu* I/*Eco* RV and, therefore, contained both introns of the *am* gene. One clone which produced a 552 base pair fragment and a 4.8kB fragment when digested with *Stu* I/*Eco* RV, suggesting a plasmid size of approximately 5.3kB, produced only a single band of approximately 2.7kB when digested with *Bam* HI. The only possible explanation for this was that the 2.7kB band represented 2 DNA fragments of equal size which were generated by digestion of the clone with *Bam*

HI. Since exactly the same banding pattern is seen when p8-3 is digested with *Bam* HI, it was decided to characterise this clone, designated pAM-8 further to determine its relationship to the genomic clone, p8-3.

4.2.6 Characterisation of pAM-8

Digestion of pAM-8 with the restriction endonuclease *Ban* II produced only two fragments of approximate sizes 5kB and 300 base pairs. Since pTZ18R contains two *Ban* II sites, at least two DNA fragments of sizes 2.3kB and 500 base pairs plus the insert size would be expected for pTZ18R containing insert. Since *am* cDNA is expected to have one *Ban* II site and *am* genomic DNA is expected to have two *Ban* II sites, more than two fragments should be generated. However, pUC-8, the host plasmid for the *am* genomic clone does not contain a *Ban* II site and the expected banding pattern for p8-3, the *am* genomic clone after digestion with *Ban* II would be two bands of sizes 292 base pairs and 5016 base pairs. This suggested that pAM-8 was actually an unaltered p8-3 clone, with the *am* genomic sequence inserted into the plasmid pUC-8 and not into pTZ18R as expected. To confirm whether this was the case, pAM-8 was digested with *Aat* II, which does not cut pTZ18R but does cut pUC-8 once and cuts the *am* gene once at position 2148. From figure 4.10A, it can be seen that uncut pAM-8 and *Aat* II digested pAM-8 (lanes 10 and 11) show exactly the same banding pattern as uncut p8-3 and *Aat* II digested p8-3 (lanes 8 and 9). Lanes 4 and 5 represent *Aat* II digested and uncut pAM-7 respectively and lanes 2 and 3 represent *Aat* II digested and uncut pTZ18R respectively. Lanes 6 and 7 represent *Aat* II digested and uncut pUC-8 respectively. It is clear from figure 4.10A that only a full length genomic *am* clone in pUC-8 would produce the banding pattern observed when digested with *Aat* II, since neither pTZ18R or pAM-7, which is missing the 3' end of the *am* genomic sequence cut with *Aat* II.

Figure 4.10B shows further confirmation that pAM-8 is identical to p8-3, since pAM-8 shows the same banding pattern as p8-3 after restriction with a variety of enzymes.

The identification of pAM-8, a clone which was isolated from the *am*-selected cDNA library, as an unmodified genomic *am* clone confirmed that the *am* selected mRNA had indeed been contaminated with p8-3. The most probable

FIGURE 4.10A

1 % agarose gel showing that digestion of pAM-8 with the restriction endonuclease, *Aat* II, produces the same restriction pattern as p8-3, which contains two *Aat* II restriction sites, one at position 2148 of the *am* sequence and one in the pUC 8 vector DNA. Since position 2148 of the *am* sequence is 3' of the mRNA transcription termination signal (see figure 4.5) and pTZ18R, the vector used for the cDNA libraries, does not contain an *Aat* II restriction site, cDNA clones to the *am* gene should not cut with *Aat* II. PAM-7, which contains 1.5kB of the 5' sequence of the*am* gene cloned in pTZ18R cannot be digested with *Aat* II.

The samples are as follows:

Track 1: 0.5 µg 1kB ladder (Gibco BRL)
Track 2: 1 µg pTZ18R digested with *Aat* II
Track 3: 1 µg undigested pTZ18R
Track 4: 1 µg pAM-7 digested with *Aat* II
Track 5: 1 µg undigested pAM-7
Track 6: 1 µg pUC 8 digested with *Aat* II
Track 7: 1 µg undigested pUC 8
Track 8: 1 µg p8-3 digested with *Aat* II
Track 9: 1 µg undigested p8-3
Track 10: 0.5 µg pAM-8 digested with *Aat* II
Track 11: 0.5 µg undigested pAM-8
Track 12: 0.5 µg 1kB ladder (Gibco BRL)

The size of the bands in tracks 8 and 10 are as shown. The marker bands in tracks 1 and 12 are as follows:

12216, 11198, 10180, 9162, 8144, 7126, 6108, 5090, 4072, 3054, 2036, 1635, 1018.

FIGURE 4.10B

1 % agarose minigel confirming that pAM-8 is identical to p8-3. Unfortunately, this figure has been printed upside down. However, the main point of this figure is to demonstrate that pAM-8 produces the banding pattern expected for p8-3, following digestion with a variety of restriction endonucleases. The track designations are correct. The samples are as follows with approximate band sizes given in brackets:

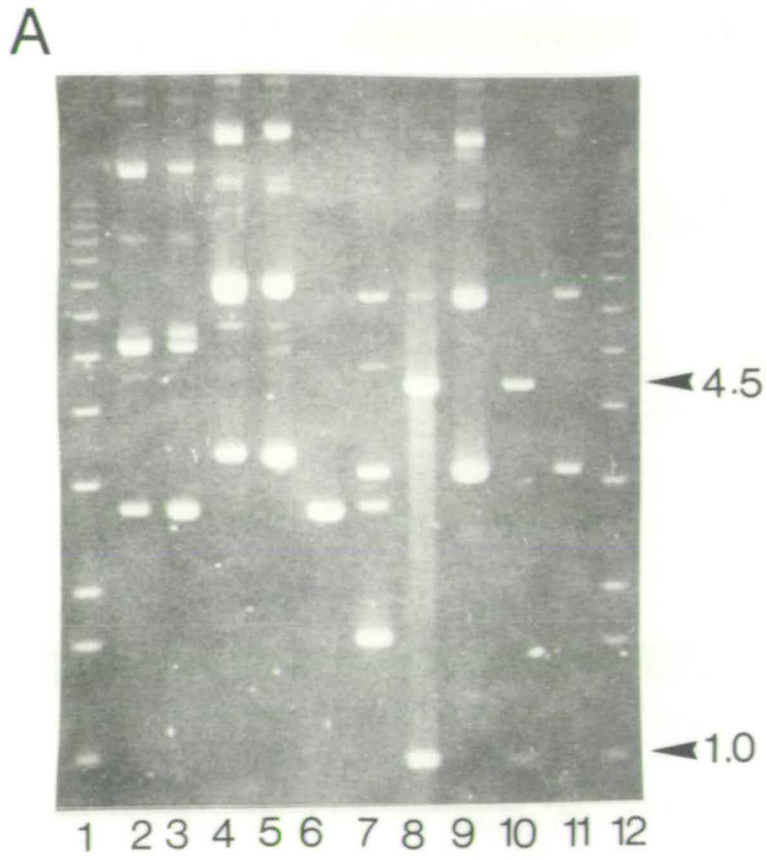
Track 1: 1 µg 1kB ladder (Gibco BRL)
Track 2: 1 µg undigested p8-3
Track 3: 0.5 µg p8-3 digested with *Hin* dIII (5.3kB)
Track 4: 0.5 µg p8-3 digested with *Bam* HI (2.7kB)
Track 5: 0.5 µg p8-3 digested with *Bgl* II (5.3kB)
Track 6: 0.5 µg p8-3 digested with *Cla* I (3.3kB, 2kB)
Track 7: 0.5 µg p8-3 digested with *Eco* RV (5.3kB)
Track 8: 0.2 µg pAM-8 digested with *Eco* RV (5.3kB)
Track 9: 0.2 µg pAM-8 digested with *Hin* dIII (5.3kB)
Track 10: 0.2 µg pAM-8 digested with *Bam* HI (2.7kB)
Track 11: 0.2 µg pAM-8 digested with *Cla* I (3.3kB, 2kB)
Track 12: 0.2 µg pAM-8 digested with *Eco* RI (3.3kB, 2kB)
Track 13: 0.5 µg undigested pAM-8
Track 14: 1 µg 1kB ladder (Gibco BRL)

The single band of 2.7kB in size seen in tracks 4 and 10 represents two DNA fragments of similar size.

The size of the marker bands are as follows:

12216, 11198, 10180, 9162, 8144, 7126, 6108, 5090, 4072, 3054, 2036, 1635, 1018, 515/506.

FIGURE 4.10



explanation for this is that some of the p8-3 which was bound to the Genescreen plus filter had been removed from the filter during the washes to recover the mRNA. Although this had been thought unlikely, it is now accepted that some DNA bound to nylon filters is removed from the filters during hybridisation and washing of Southern blots. Removal of p8-3 from the filters during the hybridisation selection procedure would also explain the higher than expected recovery of mRNA observed in sections 3.2.3 and 3.2.5.

4.2.7 Identification of the Extent of Genomic DNA contamination of the *am* -selected cDNA Library

All of the *am* positive cDNA clones which had been isolated from the *am* -selected cDNA library and characterised had contained introns. Isolating clones which produced a positive signal when probed with the 2kB *Cla* fragment of the *am* gene and using restriction digestion of the recombinant plasmid DNA to look for the presence or absence of introns was a useful method when small numbers of clones had to be screened. However 1328 clones in the *am*-selected cDNA library had produced a positive signal when probed with the 2kB *Cla* fragment of the *am* gene, so it was necessary to devise a faster method of identifying those which represented genomic *am* clones. The easiest way to do this is to use a probe specific to one of the *am* introns to probe the cDNA library. Since the primary cDNA library is stored on nitrocellulose discs from which replicas can be made, it is possible to directly compare replica filters which have been screened using different probes. A replica filter which has been screened with a probe specific for one of the *am* introns can therefore be directly compared to one which has been screened with the 2kB *Cla* fragment. Clones which produce a signal when probed with 2kB *Cla* fragment and not with the probe specific for *am* intron will either contain true *am* cDNA or will contain fragments of genomic DNA generated from the region 3' of the region homologous to the intron probe.

Due to the small size of the introns in the *am* gene, 66 and 61 base pairs, and the lack of suitable restriction sites, it is impossible to produce a restriction fragment which only covers one of the introns and no coding sequence. It was therefore necessary to construct an oligonucleotide probe homologous to one of the *am* introns. The sequence of an oligonucleotide homologous to one of the introns in

the *am* gene had to conform to the following criteria:

- it had to be as long as possible to limit cross hybridisation to other sequences;
- since the probe would be single stranded, sequences which had the potential to form secondary structure should be avoided because only linear oligonucleotide would be able to hybridise to DNA bound to a solid support such as nitrocellulose filter;
- The DNA should be as GC rich as possible as GC rich DNA has a higher melting temperature allowing high stringency washes to be used following hybridisation of the oligonucleotide probe to the target DNA sequence

These conditions significantly restricted the choice of oligonucleotide sequence. The maximum length of the intron probe was 66 bases, the length of the largest of the two introns. However, introns contain significant amounts of palindromic sequences and other sequences capable of forming secondary structure, and the introns in the *am* gene were no exception. Using the Fold programme of the UWGCG DNA sequence analysis package to identify areas of potential secondary structure, a 13 nucleotide region of the 5' intron in the *am* gene was identified as the most likely candidate for the oligonucleotide probe. The sequence of this area is GAGCTGACTTGAT and it covers positions 426-438 of the *am* gene sequence. This sequence contains 46% GC residues and covers one of the sites, GCTGACT, identified by Kinnaird and Fincham, 1983 as part of the consensus sequence which is probably essential for splicing of the intron. Using the UWGCG Fold programme to identify any potential for secondary structure formation in this sequence in isolation from the rest of the intron sequence demonstrates a lack of any such potential. The equivalent sequence of intron 2, which also contained the sequence, GCTGACT, has potential for secondary structure formation as a result of the sequence surrounding the consensus.

The sequence of the 13 base oligonucleotide is unique within the *am* gene, so would not hybridise to *am* clones which did not contain the 5' intron. To ascertain that it did not cross-hybridise to other *N.crassa* DNA sequences, *N.crassa* genomic DNA provided by Jane Kinnaird was digested with *Bam* HI and the DNA Southern blotted onto nitrocellulose filter, following gel electrophoresis of the resulting DNA fragments. Using the 13 base oligonucleotide as a probe, only a single band of 2.7kB in size, the size expected for the *am* gene, was produced. Confirmation that the clones previously isolated from the *am*-selected cDNA library did contain *am* genomic DNA was obtained

by using the 13 base oligonucleotide to probe dot blots containing plasmid DNA from these clones. pTZ18R and clones selected at random from the non-selected *N.crassa* cDNA library did not produce a signal on dot blots when probed with the 13 base oligonucleotide.

Having established that the 13 base oligonucleotide was specific for *am* genomic sequence containing the 5' intron, replicas of the filters containing the *am* -selected cDNA library were prepared as described in section 2.7 and probed with 13 base oligonucleotide, which had been end-labelled with ^{32}P using polynucleotide kinase (see section 2.14.2). The *am*-selected cDNA library was contained on four nitrocellulose filters. Of these, two showed exactly the same pattern of labelling after probing with the 13 base intron probe as they did with the 2kB *Cla* fragment of the *am* gene and with the 235 base pair *Xho* I/*Bgl* II used to establish the presence of 5' coding sequence in the clones (see section 4.2.1), indicating that all of the positive *am* clones on these filters were genomic DNA derived. The other two filters, filters 3 and 4, showed the same pattern of labelling when probed with the 13 base intron probe as they had with the 2kB *Cla* fragment. However, comparison of the filters probed with the 235 base pair fragment and the 13 base intron probe produced a slightly anomalous result. As can be seen in figure 4.11B, four weakly hybridising colonies are evident after probing filter 3 with 235 base pair probe which are not seen after probing the same filter with the 13 base intron probe (figure 4.11A). The same comparison for filter 4, shown in figure 4.12, identifies 2 weakly hybridising colonies after probing with 235 base pair probe which are not present after probing filter 4 with 13 base intron probe. It is clear in both figure 4.11 and 4.12 that the filters hybridised with 13 base intron probe (figures 4.11A and 4.12A) produced much stronger signals when exposed to autoradiography than those hybridised with 235 base pair probe (figures 4.12B and 4.12B). It is unlikely therefore that the absence of the weakly hybridising colonies from figures 4.11A and 4.12A lies in weakness of signal. Indeed overexposure of the filters did not produce signals corresponding to the colonies marked in figures 4.11B and 4.12B.

Figure 4.13 shows a comparison of filter 4 probed with 235 base pair probe (figure 4.13B) with the same filter probed with 2kB *Cla* fragment (figure 4.13C).

This comparison shows that the 2 weakly hybridising colonies identified in figure 4.12B (which is identical to figure 4.13B) are also absent in figure 4.13C. Comparison of filter 3 probed with 235 base pair probe and with 2kB *Cla* fragment also showed that the weakly hybridising colonies seen in figure 4.11B do not hybridise to 2kB *Cla* fragment. Since the weakly hybridising colonies seen after hybridisation of the 235 base pair fragment to filters 3 and 4 do not hybridise with 2kB *Cla* fragment, it is unlikely that they represent *am* clones which do not contain 5' intron as these would also be expected to produce a signal with 2kB *Cla* fragment. The most likely explanation for the observed signals is that the conditions used for hybridisation of the 235 base pair probe to filters 3 and 4 and washing of the filters after hybridisation were less stringent than those used for the 2kB *Cla* fragment and 13 base intron probe. If, for example, the temperature of the low salt washes after hybridisation with 235 base pair probe had been slightly lower than that used after hybridisation with 2kB *Cla* fragment, it is possible that DNA fragments which were not completely homologous to the 235 base pair probe retained the probe weakly bound. Unfortunately time did not permit the isolation and further characterisation of these weakly hybridising colonies from filters 3 and 4.

The absence of any detectable full length *am* cDNA clones in either the *am* -selected or non-selected *N.crassa* cDNA libraries prevented completion of the aims of the project. The production of only short *am* cDNA sequences in the non-selected cDNA library and the contamination of the *am*-selected cDNA library with p8-3, the genomic *am* clone, meant that no suitable *am* clone was available to carry out studies on the expression of *am* in *E.coli* and time limitations precluded the synthesis and screening of a new cDNA library from *N.crassa*. Repeating the cDNA synthesis with an extended 70 ^{o}C mRNA denaturation step or increasing the temperature of the denaturation step to 80 ^{o}C prior to cDNA synthesis would probably resolve the secondary structure problem which produced only short *am* cDNA clones in the non-selected *N.crassa* cDNA library, although it may be necessary to use a chemical denaturant such as methylmercuric hydroxide in the first strand cDNA reaction (Maniatis *et al.*, 1982) to guarantee denaturation of the *am* mRNA. The contamination of the *am* -selected mRNA with p8-3 was most likely a consequence of p8-3 being removed

FIGURE 4.11

Replica filters containing cDNA clones from the *am* selected cDNA library, which have been probed with (A) a 13 base pair oligonucleotide specific for the most 5' intron in the *am* gene; and (B) a 235 base pair fragment generated by digestion of the *am* gene with *Xho* I and *Bgl* II. These filters are replicas of filter 3, one of four master filters which represent the entire *am*-selected *N.crassa* cDNA library.

Arrows denote colonies which hybridise to the 235 base pair *Xho* I/*Bgl* II fragment but not to the 13 base intron probe.

FIGURE 4.12

Replica filters containing cDNA clones from the *am* selected cDNA library, which have been probed with (A) a 13 base pair oligonucleotide specific for the most 5' intron in the *am* gene; and (B) a 235 base pair fragment generated by digestion of the *am* gene with *Xho* I and *Bgl* II. These filters are replicas of filter 4, one of four master filters which represent the entire *am*-selected *N.crassa* cDNA library.

Arrows denote colonies which hybridise to the 235 base pair *Xho* I/*Bgl* II fragment but not to the 13 base intron probe.

Figure 4.13

Replica filters containing cDNA clones from the *am* selected cDNA library, which have been probed with (B) a 235 base pair fragment generated by digestion of the *am* gene with *Xho* I and *Bgl* II; and (C) a 2kB *Cla* I fragment of the *am* gene, which carries the entire coding sequence. These filters are replicas of filter 4, one of four master filters which represent the entire *am*-selected *N.crassa* cDNA library.

The filter identified as B is the same filter as B in figure 4.12.

Arrows denote colonies which hybridise to the 235 base pair *Xho* I/*Bgl* II fragment but do not hybridise to the 2kB *Cla* I fragment of the *am* gene.

Open triangles represent orientation marks on the filters.

FIGURE 4.11

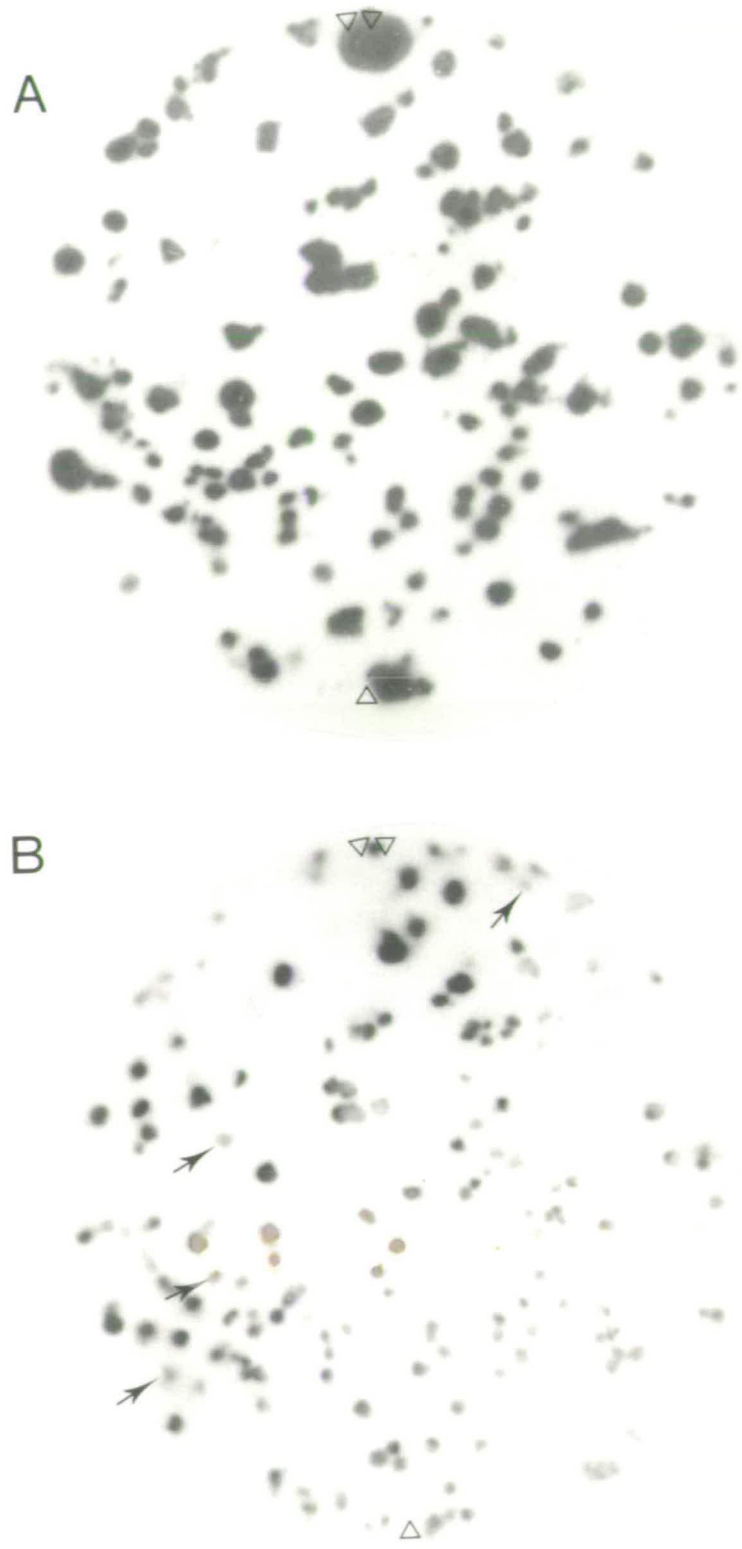
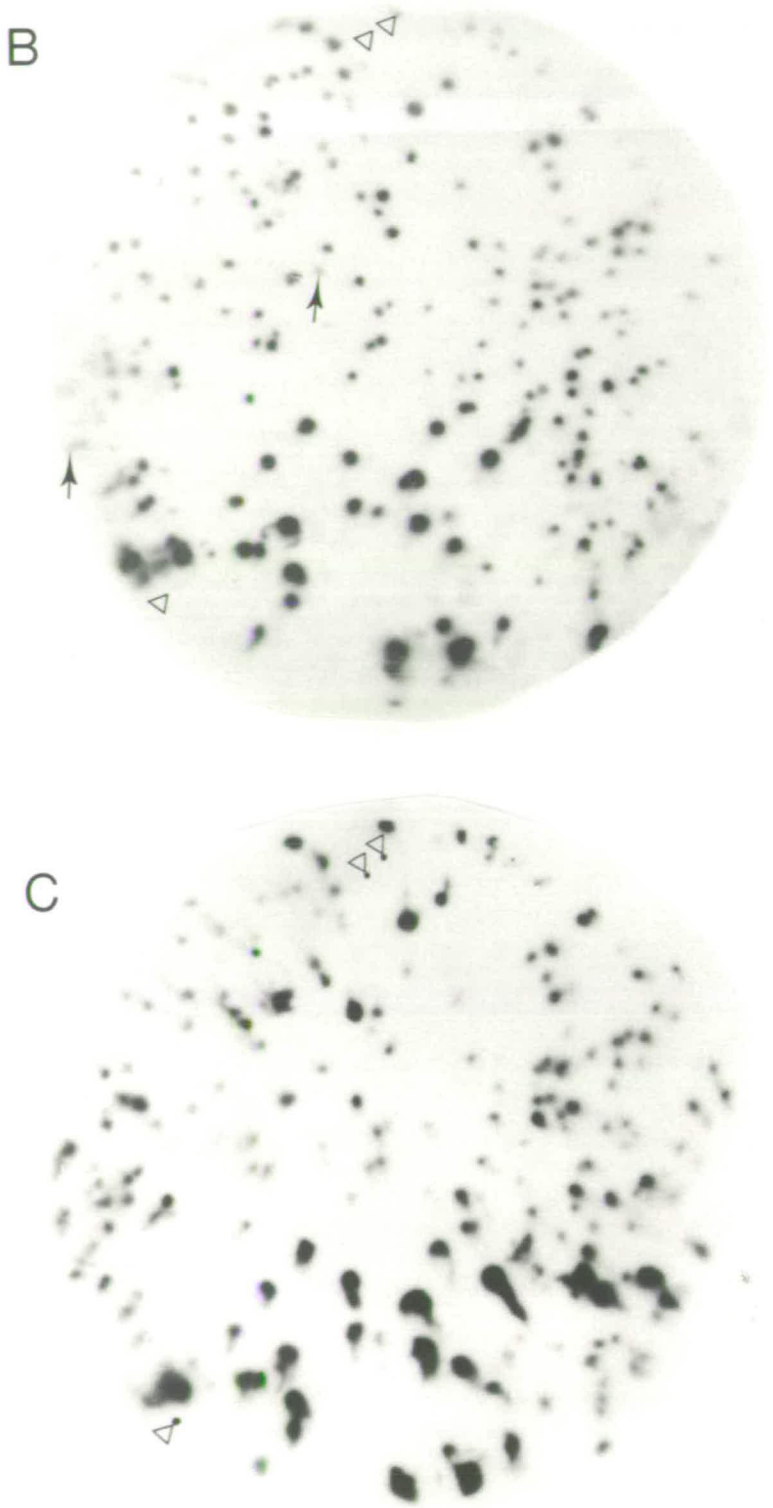


FIGURE 4.12

A

B

FIGURE 4.13



from the nylon filter during the washing procedure to recover the mRNA. Incubation of the mRNA recovered after the hybridisation selection procedure with RNase-free DNase would degrade any contaminating DNA and would have prevented the problems experienced with the *am*-selected cDNA library.

4.3 Conclusions

All of the cDNA clones identified as producing a positive signal in response to hybridisation with a 2kB *Cla* I fragment, which covers the entire coding region of the *am* gene, in the last chapter were characterised by restriction analysis or colony blotting. Of the 19 clones which produced a positive signal in the non-selected *N.crassa* cDNA library, none proved to be full length cDNA clones. The common insert size of 400-500 base pairs representing the 3' end of the *am* mRNA in all 19 clones suggested incomplete first strand synthesis by reverse transcriptase during cDNA synthesis. The presence of larger insert sizes in non-*am* derived cDNA clones suggested that the cause of this might be secondary structure in the *am* mRNA which had not been denatured by incubation at 70 ^{o}C prior to first strand synthesis. The potential to form a stable secondary structure in the area of the *am* mRNA 5' of the observed termination point of reverse transcription can be demonstrated with the help of computer programmes. This problem could be overcome either by extending the length or increasing the temperature of the 70 ^{o}C denaturing step prior to first strand cDNA synthesis; or by using a chemical denaturant such as methylmercuric hydroxide to denature the mRNA prior to cDNA synthesis.

The 1328 positive cDNA clones identified in the *am*-selected cDNA library were proved either by restriction analysis or by their ability to hybridise to an oligonucleotide probe specific to the 5' intron of the *am* gene to be derived from a genomic copy of the *am* gene. The most likely explanation for contamination of the cDNA library with *am* genomic DNA was that some of the p8-3 *am* genomic clone used to select the *am* mRNA during hybridisation selection had been removed from the nylon filter during elution of the hybridised mRNA. Confirmation of this explanation was found when one of the 'cDNA' clones, pAM-8, isolated from the *am*-selected *N.crassa* cDNA library proved to be an

exact copy of p8-3. Since p8-3 is cloned into the *Bam* HI site of pUC-8 and the cDNA library was cloned into the *Sma* I site of pTZ18R, it could be shown that pAM-8 was indeed p8-3 since it produced the restriction pattern expected from a pUC-8 clone and not a pTZ18R clone.

The absence of a full length cDNA clone for *am* from either cDNA library meant that the expression studies of *am* in *E.coli* could not be carried out.

Chapter 5 : Discussion

Chapter 5
Discussion

The major aim of this project was to develop a system which would allow the behaviour of eukaryotic DNA in response to a chemical mutagen to be studied when the DNA is resident in a prokaryotic cell and under the influence of prokaryotic repair processes. The system of choice involved the cloning and isolation of a cDNA clone for the *am* gene of *Neurospora crassa*, which could express a functional gene product in *E.coli*. It was hoped that the close similarity between the *am* gene product, NADP-dependent glutamate dehydrogenase, and its *E.coli* equivalent would allow complementation of glutamate auxotroph in *E.coli*. The major stumbling block in the project was that an appropriate cDNA clone for the *am* gene had to be produced before any further work could be undertaken. Two approaches to this were taken: one was to produce a cDNA library from *N.crassa* mRNA and screen this for a cDNA clone for the *am* gene; and the other was to produce a cDNA library specifically enriched for cDNA to the *N.crassa am* gene, by taking advantage of the homology between the *am* mRNA and the *am* genomic clone, p8-3. Both of these approaches ran into unforeseen difficulties late in the project.

The first approach produced a successful cDNA library, which contained 19 *am* cDNA clones. On analysis, the inserts in each of these clones corresponded only to a short portion of the 3' end of the *am* coding sequence. The 5' end of all of the inserts mapped to the same region of *am* DNA. The most likely explanation for this lies in the high GC content of the *am* sequence and its potential to form extensive stable secondary structure in the region 5' of the area covered by the cDNA clones. Since the *am* gene contains two introns, both of which reside in the 5' half of the gene, none of the 19 *am* clones identified in this library were suitable for expression studies. The occurrence of such a stable secondary structure in the *am* mRNA was unexpected but could have been predicted by computer analysis of the sequence. If this experiment was to be repeated, the exact melting temperature of the postulated

secondary structure would need to be calculated to ensure that the conditions used to denature the mRNA prior to cDNA synthesis were optimal for *am*. Alternatively, a chemical denaturant could be used but chemical denaturants tend to damage RNA and RNA which has been chemically denatured might not be suitable for cDNA synthesis.

The second approach attempted to specifically select mRNA for the *am* gene prior to cDNA synthesis by binding the *am* genomic clone, p8-3, to Genescreen plus nylon membrane and selecting *am* mRNA from a pool of *N.crassa* mRNA by its ability to hybridise to the bound genomic clone. The selected mRNA was then released from the bound genomic clone by heating the membrane to denature the DNA/mRNA hybrid. This approach ran into problems which were traced to the hybridisation selection procedure used to select the *am* mRNA. In addition to selected mRNA, the heat denaturation step at the end of the selection procedure also released some of the bound probe DNA from the nylon membrane. This *am* genomic DNA became represented in the cDNA library prepared from the selected material and was not detected. The reason this contamination of the cDNA library remained undetected for some time was partly due to the nature of the *am* gene and partly to the form in which the genomic DNA fragments had been released from the membrane. The genomic DNA was not represented in the cDNA library as intact *am* genomic clone, which would have been easily recognised by the size of the insert and the vector containing the DNA, which was not the same as that used for cDNA cloning, but as fragments of genomic DNA ligated into the same vector used to produce the cDNA libraries. Since a number of these clones contained inserts which were of a similar size to that expected for a cDNA clone for the *am* gene, it was assumed that they were cDNA clones. Early restriction mapping appeared to confirm this. The introns in the *am* gene are both very short (61 and 66 base pairs) and their presence cannot easily be identified by restriction mapping as there are no unique restriction sites in the introns. The size difference between predicted restriction fragments of *am* cDNA and those of *am* genomic DNA containing one or more introns is very small so early restriction mapping studies could not detect their presence in the putative cDNA

clones. It was only when the restriction map of the 5' end of the cDNA insert demonstrated a high level of similarity to the 5' end of the genomic DNA that the insert was recognised as genomic in origin. The use of an oligonucleotide probe specific for one of the introns of the *am* gene proved that all of the clones in the cDNA library containing *am* DNA were derived from genomic DNA and not cDNA. Had this probe been available early in the project, an initial screen of the cDNA library using it would have identified the problem at the outset. Unfortunately, the probe was not available until late in the project and could only be used to confirm what had already become apparent from extensive restriction mapping of the clones.

In retrospect, the earliest evidence that there may have been a problem with the hybridisation selected material came during the cDNA synthesis itself. Considerably more first strand cDNA was synthesised from the *am*-selected material than would have been expected (see section 3.2.5) if only *am* mRNA had been selected during hybridisation selection. Two explanations were identified for this: either *am* mRNA constituted more than 0.1% of *N.crassa* mRNA; or other RNA species had also been selected in the procedure. The latter explanation was thought more likely for several reasons. Firstly, the proportion of *am* cDNA clones in the cDNA library made using non-selected *N.crassa* mRNA was 0.11%, which agreed with the figure quoted for representation of *am* mRNA in *N.crassa*. Secondly, there was already evidence that the *am* genomic clone, p8-3, was capable of hybridising to other *N.crassa* RNA species under conditions of high stringency (see p.86 and figure 3.4) and, since the conditions used for the hybridisation selection were less stringent, these RNA species would almost certainly have been represented in the selected material. A third reason for believing that *am* mRNA was not the only species isolated by the hybridisation selection procedure was that the number of recombinants in the *am*-selected cDNA library which contained *am* sequence only constituted 10% of the total recombinants in the library, suggesting that the other 90% were derived from non-*am* sequences. With the knowledge that genomic DNA had co-eluted from the membrane during the hybridisation selection procedure, it can now be argued that the clones which do not

contain *am* sequence probably contain inserts which are derived from areas of the *am* genomic clone, p8-3, which were not covered by the 2kB *Cla* I fragment used to probe the library. If this is the explanation, it is interesting to note that the number of clones containing DNA derived from p8-3 which is homologous to the 2kB *Cla* I fragment should have constituted much greater than 10% of the library, assuming that each area of the plasmid was equally represented and given the total plasmid size of 5.5kB.

If DNA was responsible for the larger amount of cDNA synthesised than was expected, then it must have been copied by reverse transcriptase during the first strand cDNA synthesis because ^{32}P-dCTP was incorporated into DNA during this reaction. Indeed, the amount of first strand cDNA synthesised was calculated from the incorporation of ^{32}P-dCTP. Although reverse transcriptase is known to have DNA-dependent DNA polymerase activity, the conditions used for cDNA synthesis should select against this activity so, in theory, DNA should be copied inefficiently. Assuming that co-eluted probe DNA is the reason for the greater than expected yield of cDNA, then the first strand cDNA synthesis reaction conditions did not select against the DNA-dependent DNA polymerase activity of reverse transcriptase. If this is the explanation, then the copied DNA fragments, or other single stranded DNA fragments present in the mix, must have acted as templates for *E.coli* DNA Polymerase I in the second strand reaction as ^{32}P-dCTP was also incorporated into the DNA in this reaction at the level normally expected for a second strand cDNA reaction. It is possible that small fragments of eluted DNA acted as primers in the reactions with larger fragments of eluted DNA acting as template either for the DNA-dependent DNA polymerase activity of reverse transcriptase in the first strand synthesis reaction or for *E.coli* DNA Polymerase I in the second strand synthesis reaction. The resulting double stranded DNA would have been made blunt ended by T4 DNA Polymerase, thus producing ends compatible for ligation into *Sma* I digested pTZ18R. This is the only explanation which accounts for the incorporation of ^{32}P-dCTP into DNA synthesised in both first and second strand reactions and for the

appearance of *am* genomic DNA fragments in pTZ18R, the vector used for cDNA cloning. Because of the number and complexity of reactions involved, the only way to control for this would have been to:

1. perform the hybridisation selection in the absence of RNA;
2. use the eluate from the hybridisation selection as a template for cDNA synthesis, and ligate the product into pTZ18R;
3. transform the ligation products into competent *E.coli* cells.

Even if this procedure is followed, there is no way to ensure that each reaction in the complex series required would behave identically to those undergone by the eluate used to produce the *am*-selected cDNA library. It could be argued that simply demonstrating that probe DNA had eluted from the filter would have been sufficient to prevent the problems encountered in this thesis. This could be done by performing a hybridisation selection in the absence of RNA and using the eluate in a dot blot such as that shown in figure 3.4. However, the next logical step in this procedure would have been to transform the eluted DNA into *E.coli* to determine the contribution it would make to the cDNA library. This would have shown that the contribution of directly eluted p8-3 to the cDNA library was less than 0.1%, since clones identical to p8-3 only constitute 0.66% of the library, and so would not have been considered significant. Since the major problem encountered in this work was not elution of DNA from the membrane, but elution of DNA fragments which were either clonable or capable of becoming clonable into *Sma* I digested pTZ18R, simply demonstrating that DNA elutes from nylon membrane would not have been sufficient.

By far the easiest remedy to the problems encountered would be to treat the hybridisation selected material with RNase-free DNase following elution from the membrane to degrade any contaminating DNA. This would ensure that only RNA was present in the cDNA synthesis reaction. Had the potential for DNA elution from nylon membranes been realised at the time this work was done, this is a step I would have undoubtedly included in the protocol.

At the time this hybridisation selection experiment was done, a major stumbling block was a lack of information regarding elution of bound DNA from nylon membranes. All of the available literature stated that DNA, once bound to the membrane by baking or UV crosslinking, was irreversibly bound and could not be recovered from the membrane. There were, and, indeed, are still, no publications which refer to this happening.

At the time this work was done, hybridisation selection using DNA bound to a solid support such as nylon or nitrocellulose membrane as a means to select a particular nucleic acid species for further manipulation was considered a novel procedure. The original references describing hybridisation selection (Parnes *et al.*, 1981; Ricciardi *et al.*, 1979) used DNA bound to nitrocellulose membrane to select specific mRNA species for *in vitro* translation. Neither method took any steps to remove DNA which may have eluted from the membrane, despite the fact that traditional *in vitro* translation systems require the removal of DNA for efficient translation. Since the mRNA selected in both of these papers was suitable for *in vitro* translation, there was no reason to suppose that mRNA selected in this way would not be suitable for other mRNA manipulation procedures such as cDNA synthesis.

More recently, hybridisation selection has found increasing application in the field of molecular biology. It has been used as a means to select specific cDNA and genomic clones from plasmid libraries following single stranded rescue of the entire library (Love & Deininger, 1991[1]; Pruitt, 1988[1]); to select genomic DNA fragments prior to cloning (Colgan, 1988[1]); and to select a mixture of cDNA species which are homologous to a cosmid or YAC genomic clone for amplification by the polymerase chain reaction (Saiki *et al.*, 1988) and cloning into vector to produce a cDNA library specifically enriched for the coding sequences contained on the genomic clone (Lovett *et al.*, 1991[1]; Parimoo *et al.*, 1991[1]). Only one of these papers (Lovett *et al.*, 1991[1])

---

[1] See Additional References section.

contains a control to ascertain whether DNA which may have eluted from the membrane is represented in the subsequent library. In at least two of these papers, those by Colgan[1], 1988, who used Genescreen plus nylon membrane, and by Parimoo *et al.*, 1991[1], who used Hybond (Amersham) nylon membrane, there is evidence that there may be contamination in the resulting libraries derived from elution of the original probe DNA from the membrane. Lovett *et al.*, 1991[1] did include a control to identify contamination of the resulting library derived from eluted probe DNA. This consisted of amplifying a control eluate from a filter containing a YAC clone which had not been hybridised with cDNA, using polymerase chain reaction. In this case, they did not find any detectable contamination (although they do report false positive clones in the library produced using the eluate from filters containing YAC clones which had been hybridised with cDNA).

For many applications involving hybridisation selection, co-elution of the probe DNA with the selected nucleic acid may not create a problem, if, for example, the probe DNA is linear and the selected material is plasmid DNA which will be directly transformed into bacteria (Love & Deininger, 1991[1], Pruitt, 1988[1]), or if the probe DNA has ends which are unsuitable for ligation with the vector used to clone the selected DNA (Colgan, 1988[1]). However, procedures which involve DNA synthesis following selection such as cDNA synthesis or amplification by polymerase chain reaction are likely to produce artefactual results as a result of co-elution of probe DNA with the selected nucleic acid, so the knowledge that this can happen is particularly important in these applications. Obviously, where the selected material is RNA, the introduction of an RNase-free DNase digestion step is all that is required following hybridisation selection to prevent the problems reported in this thesis. For procedures which select DNA, especially those which require amplification of the DNA by polymerase chain reaction, it is not possible to easily remove the probe DNA from the selected DNA following hybridisation selection. Although it is

[1] See Additional References section

possible to include a control reaction such as that used by Lovett *et al.*, 1991[1], this still does not ensure that no probe DNA has contaminated the sample DNA used to produce the library of interest. By far the safest way to proceed using hybridisation selection for these procedures is to use an alternative support which will not allow the probe DNA to elute. A much more efficient method for hybridisation selection has recently been developed for carrying out a similar procedure to that described by Lovett *et al.*, 1991[1] and Parimoo *et al.*, 1991[1] by Finbar Cotter at the Institute of Child Health in London (personal communication). This method involves labelling the genomic probe DNA using biotin, allowing this to hybridise to a mixture of cDNA species in solution, then binding the genomic DNA/cDNA hybrids to streptavidin coated paramagnetic particles by taking advantage of the ability of biotin to bind irreversibly to streptavidin. The DNA hybrids which are bound to the particles can be separated from the non-selected material by the application of a magnetic field, and, following washing, the selected material can be eluted by heating, leaving the probe DNA firmly bound to the paramagnetic particles by virtue of the biotin/streptavidin link. This method produces very pure selected material which can be used in polymerase chain reaction, in the knowledge that it is not likely to be contaminated with genomic probe DNA. The solid support for this particular method was chosen after careful consideration of the problems reported in this thesis, which are typical when using nylon membrane for this kind of procedure. Paramagnetic particle technology is almost certainly the safest way to undertake hybridisation selection procedures and, because of the versatility of the system, it can be adapted for any application which requires selection of particular RNA or DNA species for subsequent manipulation and cloning. Had this technology been available at the time the work reported in this thesis was carried out, it would have been the ideal system to use in selecting the mRNA for the *am* gene. Since this work was done there have been a number of other new technologies developed which would have been invaluable to this project. These are discussed in the next section.

---

[1] See additional References section

5' of the oligo dT. Ligation of adaptors to the cDNA molecules, to produce a second restriction site, followed by digestion at the site contained within the oligo dT primer, produces cDNA molecules with non-compatible ends, which can then be ligated into the appropriately digested vector DNA. The choice of restriction sites dictates the orientation of the cDNA molecules in the vector.

The pBluescript plasmid contains all of the advantages offered by pTZ18R for cloning, including blue/white screening for recombinants, single stranded rescue for sequencing, the ability to produce fusion proteins, and a T7 promoter for *in vitro* transcription. cDNA clones, isolated by single stranded excision of pBluescript from Lambda ZAP, can therefore be dealt with in the same manner as clones in pTZ18R. In addition, the pBluescript plasmid has a much larger polylinker than pTZ18R and contains a T3 promoter, in the opposite orientation from the T7 promoter, which can also be used for *in vitro* transcription, so is a more versatile vector than pTZ18R. The use of Lambda ZAP as the vector for a *Neurospora crassa* cDNA library in this study would have allowed the construction of a much larger cDNA library than those produced, and the production of a cDNA library in which all the cDNA molecules were inserted into the *LacZ'* gene in the correct orientation for expression. Following identification of *am* positive cDNA clones, the pBluescript plasmids containing the cDNA could be excised from the lambda DNA with the aid of a helper phage and either transformed directly into PA340 and tested for their ability to complement glutamate auxotrophy in *E.coli*, or characterised in the same manner as the pTZ18R clones in this study. The main advantage offered by vectors such as Lambda ZAP is the ability to produce a high titre directional cDNA library which has all the versatility associated with plasmid vectors. The production of a high titre cDNA library obviously increases the possibility of obtaining a full length clone for the cDNA of interest. The use of directional cDNA cloning reduces by a factor of two the number of cDNA clones which must be screened to identify a clone is capable of expressing the cDNA of interest. These factors coupled with the ability to produce a plasmid cDNA clone from the lambda clone without the need to subclone and the versatility of the pBluescript plasmid would have significantly affected the choice of cloning strategy used in this study, had such vectors been available when this work was

done. However, the use of a more advanced vector system, would not have addressed the problems associated with isolation and characterisation of potential *am* cDNA clones which were discussed in the last chapter.

A recent development which could have proved invaluable to this project is the use of the polymerase chain reaction to produce cDNA molecules of specific sequence (Frohman *et al.*, 1988). This system of RACE PCR (Rapid Amplification of Complementary DNAs by Polymerase Chain Reaction) takes advantage of the poly(A) tail of mRNA molecules and the ability of *Taq* DNA polymerase to undergo thermal cycling and amplify specific DNA sequences in the presence of appropriate primers in the polymerase chain reaction (Saiki *et al.*, 1988). Using RACE PCR, first strand cDNA is prepared using reverse transcriptase and oligo dT primer. The reverse transcriptase can then be denatured by heating to 70 ^{o}C for ten minutes and the *Taq* polymerase and primers added. The primers in RACE PCR are normally oligo dT and a primer specific for the cDNA sequence of interest. The sequence of interest can then be amplified by several rounds of denaturation, annealing and elongation. This system requires that some of the sequence at the 5' end of the mRNA is known to allow the design of an oligonucleotide specific for the 5' end of the mRNA. The *am* mRNA would be ideal for this as the genomic sequence is known. It would also be possible to take advantage of the gene sequence to make the reaction more specific for amplification of *am* cDNA by constructing oligonucleotides specific for the 5' and 3' ends of the *am* mRNA and using these as primers to amplify the *am* coding sequence by PCR. The use of specific primers to amplify cDNA sequences by PCR following first strand cDNA synthesis by reverse transcriptase has been described by Gama *et al.*, 1988 and Kawasaki, *et al.*, 1988. The choice of appropriate primers should minimise extraneous non-coding sequence and ensure that cDNA for the *am* gene is the major product of the amplification reaction. For example, the TAA in frame stop codon, which is generated by expressing *am* as a fusion protein with the β-galactosidase alpha peptide, could be omitted from the *am* sequence without the loss of coding sequence by choosing a primer which is homologous to part of the *am* sequence 3' of the TAA stop codon and 5' of the ATG start codon of the *am* gene. This would increase the possibility of expressing a functional *am* gene

product in *E.coli*. Unfortunately *Taq* DNA polymerase contains an exonuclease activity which can 'nibble' the ends of the PCR product, so the frame of the cDNA could not be guaranteed after cloning. However, the inclusion of a restriction endonuclease recognition sequence in the primers would allow insertion of the PCR product into a specific cloning site on the plasmid polylinker. It should therefore be possible to produce cDNA clones in which the *am* coding sequence is in frame with the *LacZ'* coding sequence by choosing an appropriate restriction site on the plasmid. The use of different restriction endonuclease recognition sites on each primer and appropriately digested plasmid DNA would provide a means to ensure that the *am* cDNA is inserted into the plasmid in the correct orientation for expression as well as in the correct reading frame with respect to the *LacZ'* coding sequence.

PCR is a powerful new technique which can be used as a short cut towards the production of in frame cDNA clones which can be expressed in *E.coli*. If PCR technology had been available at the time when this project was undertaken, the approach to producing an *am* cDNA clone for expression in *E.coli* would have been very different from the one undertaken. The production of an in frame cDNA clone for the *am* gene in a plasmid vector using PCR would have significantly reduced the amount of time spent on generating cDNA libraries and characterising potential *am* cDNA clones during the course of this work and removed much of the uncertainty which is an integral part of these procedures. This would have allowed work to progress on the expression of *am*-encoded NADP-GDH in *E.coli* and the introduction and characterisation of chemically induced frameshift mutations in the *am* sequence.

Although the work described in this thesis did not fulfil the initial aim of the project, which was to set up a system which would allow identification and characterisation of frameshift mutations induced in a eukaryotic gene which is resident within a prokaryotic environment, it is interesting to look at similar work which has been published since this work was done.

Stanolowski *et al.*, 1986, published a piece of work which complements what this work set out to do. They used the chemical mutagen ICR 191 to induce mutations in an *E.coli* gene which was stably integrated into the genome of a

mammalian cell line. They used the Chinese hamster ovary cell line, AS52, which contains a functional copy of the *E.coli* gene for xanthine-guanine phosphoribosyl transferase (XPRT), *gpt*. AS52 is deficient in hypoxanthine-guanine phosphoribosyl transferase (HPRT). Since the spontaneous mutation rate at the *gpt* locus in AS52 cells compares favourably with the spontaneous mutation rate at the *hprt* locus of wild-type cells, mutations induced at each loci can be compared directly. Treatment of the wild type and AS52 cell lines with ICR 191 showed that the *E.coli gpt* gene was more susceptible to mutation by ICR 191 than the Chinese hamster gene *hprt*. The ratio of XPRT to HPRT mutants induced per unit dose of ICR 191 was 1.6. The same ratio for XPRT and HPRT mutants induced with ethyl methanesulfonate (EMS) is 0.7, inferring that the observed ratios of induced XPRT to HPRT mutations are a function of the mutagen used. Unfortunately, Stanolowski *et al.* only characterised mutations which changed the restriction pattern of the genes and no information was presented about the nature of the mutations induced or their position within the two genes. The higher mutation rate demonstrated by the *E.coli gpt* gene in response to ICR 191 suggests that sequence specificity plays an important role in induction of mutations by ICR 191. However, no sequence information was given about the *gpt* or *hprt* genes or the induced mutations, so no conclusions can be drawn regarding the nature of the sequence specificity. Characterisation of the mutations induced in the XPRT and HPRT genes by ICR 191 in this study would provide an invaluable contribution to understanding the mechanism by which frameshift mutations are induced in prokaryotic and eukaryotic DNA by acridine half-mustards.

As stated in the Introduction to this thesis, ICR 191 is more active when used to induce mutations in prokaryotic DNA than eukaryotic DNA, whereas ICR 170 is more active when used to induce mutations in eukaryotic DNA. One of the concerns about using ICR 170 to induce mutations in eukaryotic DNA which is resident in a prokaryotic environment or ICR 191 to induce mutations in prokaryotic DNA resident in a eukaryotic environment is whether the mutagen will be sufficiently active to produce mutations in the gene of interest. The induction of mutations by ICR 191 in the *E.coli gpt* gene demonstrated by Stanolowski *et al.* shows that ICR 191 is capable of inducing mutations in DNA

contained within eukaryotic cells but can ICR 170 induce mutations in DNA in prokaryotic cells? A study by Hoffmann *et al.*, 1989, into the induction of genetic duplications and frameshift mutations by acridines and acridine mustards in *Salmonella typhimurium* provides an answer. Comparison of the induction of frameshift mutations in *hisC3076*, an allele which contains a mutation which can be reverted by frameshift mutagens to produce a His^+ phenotype, by ICR 191 and ICR 170 shows that, although a higher frequency of reversion to His^+ is seen when TS869, an *S.typhimurium* strain which is deficient in excision repair, is treated with ICR 191 than with ICR 170, the opposite is true of TS1125, an *S.typhimurium* strain which has a functional excision repair pathway. It is clear, therefore, that ICR 170 is capable of inducing mutations in prokaryotes.

Hoffmann *et al.*, 1989, also show that acridine compounds which intercalate into DNA but do not form covalent attachments to the DNA are capable of producing frameshift mutations but not duplications. Conversely, compounds which can form covalent adducts with the DNA but do not intercalate can form both duplications and frameshift mutations. Both types of compound produce frameshift mutations at lower frequencies than ICR 191, ICR 170 or the related compound, ICR 372, which intercalate into the DNA and bind covalently to it (DeMarini *et al.*, 1984). This suggests that, although simple intercalation of the mutagen into, or covalent binding of the mutagen to the DNA can produce frameshift mutations, compounds which can do both are more potent mutagens, at least in *S.typhimurium*. The suggestion by Streisinger and Owen, 1985, that the mutagen molecules may 'stack' on either side of a misaligned base, thus stabilising it, may not be far from the truth. It can be envisaged that intercalation of a mutagen molecule opposite a misaligned base, coupled with covalent binding of the mutagen to the DNA adjacent to the misaligned base would be more likely to stabilise the misalignment and result in a frameshift mutation than intercalation or covalent binding of the mutagen alone. Sequence specificity of the mutagen for the formation of covalent bonds with the DNA may, in part, account for the occurence of mutation 'hotspots' in response to induction by frameshift mutagens. Another mutagen which is known to produce predominately frameshift mutations and which attaches covalently to the DNA is the carcinogen, N-2-acetylaminofluorene (AAF). Using a related chemical,

N-acetoxy-N-2-aminofluorene (N-aco-AAF) to induce mutations in the tetracycline resistance gene of the plasmid, pBR322, Koffel-Schwartz *et al.*, 1984, 1987 found that 90 % of the mutations induced were located at specific hotspots. They defined two classes of mutation hotspot in the tetracycline resistance gene:

1. homopolymeric stretches of guanine residues, which produce -1 frameshift mutations after induction with N-aco-AAF (i.e. GGGG-->GGG)
2. short stretches of alternating GC residues, such as the *Nar* I restriction site, which produce -2 frameshift mutations after induction with N-aco-AAF (i.e. GGCGCC-->GGCC)

Koffel-Schwartz and Fuchs, 1989, suggest that the first class of mutation, -1 frameshift mutations at runs of guanine residues, arises by the misaligned pairing model suggested by Streisinger and Owen, 1985, and that the conformational change in the DNA induced by covalent binding of N-aco-AAF stabilises the misaligned base. N-aco-AAF binds covalently to position C-8 of guanine and causes the guanine residue to rotate from the *anti* to the *syn* conformation, resulting in insertion of the fluorene ring of N-aco- AAF in the DNA double helix, (Fuchs *et al.*,1976). Perhaps a similar explanation exists for the induction of frameshift mutations by the acridine half-mustards. N-aco-AAF is also capable of converting B DNA to Z DNA when it binds to synthetic polynucleotides containing alternating purine-pyrimidine sequences (Sage and Leng, 1980, 1981) and a different explanation for the second class of mutation (-2 frameshifts at alternating GC sequences) is proposed. This suggests that this class of mutation occurs as a result of a pathway designated the '*Nar* I mutation pathway'. This pathway requires the expression of one or more of the gene products involved in the SOS repair pathway and activation of the RecA protein. Although the role of unidentified SOS gene products has not been elucidated, Koffel-Schwartz and Fuchs, 1989, do suggest a role for the RecA protein. Based on the observation that the inactivated form of the RecA protein inhibits the *Nar* I mutation pathway, they have suggested that RecA protein binds directly to the N-aco-AAF modified DNA target site and stops the formation of a mutation. Activation of the RecA protein abolishes its ability to bind to the target site and mutagenesis can occur. Confirmation that two different pathways are involved in the formation of N-aco-AAF induced mutations comes from the observation that the induction of -1 frameshifts at runs of guanine is *umuDC* dependent, whereas

the induction of -2 frameshifts at alternating GC residues is *umuDC* independent. No explanation was offered as to why N-aco-AAF should only induce frameshift deletions in the DNA.

The acridine half-mustards predominately produce single base frameshift mutations at monotonous runs of G or C residues, so it is likely that they conform to the misaligned pairing model postulated by Streisinger and Owen, 1985 rather than utilising the *Nar* I pathway to induce mutations.

Hsieh and Griffith, 1989, demonstrated that deletion of 1-4 bases in DNA produces a sharp 'kink' in the DNA molecule which can be detected by gel retardation assay and NMR studies, suggesting that the base(s) opposite the deletion is forced out of its position in the DNA helix. DNA containing single base mismatches does not produce any abnormalities in the DNA structure which could be determined using these techniques. A similar study by Bhattacharyya and Lilley, 1989 , describes the local DNA structure at the position opposite a deleted base as a 'base bulge' in the DNA helix and suggests that the bulge results in a stable distortion of the DNA as a result of the necessity to accommodate extra-helical nucleotides. Variation in chemical reactivity of bases involved in the looped out portion of the base bulge was also observed, with bases at the 5' end of the bulge less susceptible to chemical attack than those at the 3' end. It was suggested that the bases at the 5' end may be involved in stacking reactions with the duplex DNA. The well defined structure of base bulges may make them easily recogniseable to proteins of the repair systems and explain why spontaneous deletion mutations are favoured *in vivo* (see Introduction). It would be very interesting to see what effect the addition of a frameshift mutagen to the DNA molecule has on its behaviour in a gel retardation assay. NMR and crystallography techniques may provide useful information about the interaction of the mutagen molecule with the DNA. It is possible that the mutagen molecule may somehow shield the bulge from protein recognition or that, by altering the conformation of the DNA in a similar fashion to that described for N-aco-AAF, the mutagen prevents the formation of recogniseable bulges during the formation of frameshift mutations.

The description of base bulges as intermediates in the formation of frameshift

mutations appears to support the misaligned pairing model suggested by Streisinger and Owen, 1985.

Since this work was done, polymerase specific hotspots for spontaneous frameshift mutations have been identified. De Boer and Ripley, 1988 identified a consensus sequence for -1 frameshift mutations which are produced after DNA synthesis *in vitro* by the Klenow fragment of *E.coli* polymerase I, and which do not conform to the Streisinger and Owen model of misaligned pairing. Another interesting observation they made was that DNA polymerised *in vitro* by Klenow fragment produced only deletion mutations instead of the additions and deletions which are normally seen when DNA is polymerised *in vivo* by this enzyme. Papanicolaou and Ripley, 1989, observed that different site specificities for the formation of complex frameshift mutations were demonstrated by *E.coli* polymerase I holoenzyme and the Klenow fragment of *E.coli* polymerase I, although both demonstrate sensitivity for the consensus sequence involved in the production of -1 frameshift mutations.

Polymerase specific hotspots for frameshift mutation have also been demonstrated in the bacteriophage T4 (Ripley *et al.*, 1983).

New assay systems are now available for studying frameshift mutagenesis. Koehl *et al.*, 1989, developed a plasmid system for studying the induction of frameshift mutations at alternating GC residues by AAF and N-aco-AAF. This assay system utilises the *lacZ* alpha complementation assay in *E.coli*. By introducing 14 bases, carrying a *Nar* I restriction site, into the *lacZ'* region of a plasmid, pEMBL 8, Koehl *et al.* not only inserted a unique *Nar* I restriction site in the plasmid, but also introduced a +2 frameshift into the *lacZ'* gene. Bacteria containing this plasmid therefore appear white on plates containing X-gal. The introduction of mutations by AAF or its derivatives via the *Nar* I mutation pathway result in a two base deletion at the *Nar* I site and restore the reading frame of the *lacZ'*. This is a variation of the system described as System 2 in the Introduction to this thesis. The system described by Koehl *et al.* is specific for mutations induced via the *Nar* I mutation pathway and will not detect single base deletions. However, it is possible single base additions induced in this sequence would be identified as these would also restore the reading frame of *lacZ'*, so

the system may have more widespread uses.

Other frameshift assay systems developed recently have also been based on variations of System 2. De Boer and Ripley, 1988, developed a system for studying single base pair deletions induced by the Klenow fragment of *E.coli* polymerase I. They introduced 77 bases of the T4 *rIIB* DNA into M13 in such a way as to produce two clones, P74, carrying a -1 frameshift in the *lacZ'* gene of M13 and M74, containing a +1 frameshift in the *lacZ'* gene of M13. The introduction of a single base addition into P74 or a single base deletion into M74 results in restoration of the alpha complementation function of M13 and the production of blue plaques in the presence of X-gal. The main difference between this system and the one described in the Introduction to this thesis is that the insert DNA renders the *lacZ'* gene out of frame in this system, and the introduction of frameshift mutations subsequently restores the *lacZ'* reading frame and its ability to produce a functional β-galactosidase alpha peptide. System 2, described in the Introduction to this thesis, required insertion of the insert DNA in a way which would not disrupt the reading frame of *lacZ'* and the identification of mutants by loss of β-galactosidase activity. The system described by De Boer and Ripley, 1988, is a particularly nice one as the nature of the frameshift, i.e. whether it is an addition or deletion mutation, is immediately apparent.

A similar approach was used by Cronan Jr. *et al.* in 1988 to produce vectors which could be used to assemble synthetic genes. This system involved producing frameshifts in the *lacZ'* gene of PUC 18 and pTZ18R plasmids. Having constructed +1 and +2 frameshifts in the plasmids, Cronan Jr. *et al.* introduced synthetic DNA molecules 60-90 base pairs in length and used restoration of β-galactosidase activity as an indication of in frame fusion of the synthetic DNA. This would have been a nice system to use to produce in frame fusions of *am* DNA for use in the mutagenesis studies described for System 2. It is interesting to note that the DNA inserts used by De Boer and Ripley and Cronan Jr. *et al.* were similar in size so it may be that this represents a practical limit to the size of DNA fragment which can be inserted into the *lacZ'* gene without losing activity.

The advent of new technologies and novel assay systems for frameshift mutations make it possible to generate information at the DNA sequence level about the response of DNA to induction of mutations by chemicals. It is now possible to produce specific cDNA sequences using the polymerase chain reaction, without the need to construct and screen cDNA libraries. The use of insertional restoration of β-galactosidase as described by Cronan Jr. *et al.*, 1988, provides an ideal way to produce *am* sequences suitable for mutagenesis and screening by the method described as System 2 in the Introduction to this thesis. However, the concern about whether small pieces of *am*, mutated in isolation from the rest of the gene, would be representative of the behaviour of the *am* gene in response to frameshift mutagens *in vivo* is still valid. It would always be necessary to look at the behaviour of the intact gene in response to treatment by chemical mutagens so that the structure of the DNA is considered. There is no reason to prevent the approaches described as System 1 and System 2 being used in parallel. Indeed, this may be an ideal way to generate information. The blue/white frameshift assay described for System 2 would provide a quick method of generating mutants which could then be characterised by sequence analysis, and mutations generated using System 1 would provide more comprehensive information about the response of the *am* gene to the frameshift mutagen ICR 170.

Although much work remains to be done if the mechanisms responsible for frameshift mutagenesis are to be understood, it is likely that the developments discussed in this chapter will play an important part in future studies.

Appendix A : Materials

Appendix A
Materials

A.1 Bacterial Strain Genotypes

NM522

supE, thi, Δ(lac-proAB), Δ hsd5 (R^-M^-), [F' proAB $lacI^qZ\Delta M15$]

PA340

thr-1, ara-13, tonA2, leuB6, lacY1, supE44, gal-6, λ^-, gdh-1, hisG1, rfbD1, galP63, gltB31, rpsL9, malA1(λ^R), xyl-7, mtl-2, argH1, thi-1

A.2 Oligonucleotide probes

Oligonucleotide probes were obtained from OSWEL DNA Service, University of Edinburgh.

A.3 Media

Vogel's Minimal Salts (50 X), 1 litre

150 g Na_3 citrate
250 g KH_2PO_4
100 g NH_4NO_3
10 g $MgSO_4.7H_2O$
5 g $CaCl_2$
5 ml Trace elements
1-2 ml Chloroform
12.5 ml Biotin stock (1 % w/v in 50 % ethanol)

This was made up to 1 litre with distilled water and sterilised by autoclaving.

Vogel's Trace Elements, 100ml

5 g Citric acid
5 g $ZnSO_4.7H_2O$
1 g $Fe(NH_4)_2(SO_4)_2.6H_2O$
0.25 g $CuSO_4.5H_2O$
0.05 g $MnSO_4.H_2O$
0.05 g H_3BO_3
0.05 g $NaMoO_4.2H_2O$

This was made up to 100 ml with distilled water.

Vogel's Liquid Medium 1 X Vogel's Minimal Salts
1.5 % Sucrose

Vogel's Agar Slants 1 X Vogel's Minimal Salts
1.5 % Sucrose
1 % Agar

This was dispensed into large test-tubes (5 ml/tube]. The tubes were plugged with cotton wool and sterilised by autoclaving at 121 ^{o}C, 15 lb/in^2 for 15 minutes. After autoclaving, the tubes were placed at an angle and allowed to set at room temperature.

M9 Minimal Salts (10 X) 6 % NaH_2PO_4
3 % KH_2PO_4
0.5 % NaCl
1 % NH_4Cl

M9N Minimal Salts (10 X) 6 % NaH_2PO_4
3 % KH_2PO_4
0.5 % NaCl
2 % NH_4Cl

M9 Minimal Medium 1 X M9 salts
1 mM $MgSO_4$
0.2 % glucose
1 % thiamine
0.1 mM $CaCl_2$

All solutions were autoclaved separately as 10 X stocks, except thiamine, which was sterile filtered, and then added to sterile flasks immediately prior to use. For M9 agar, the solutions were added to a sterile 2 % agar solution at approximately 50 ^{o}C before the agar was poured into petri dishes and allowed to set.

M9N Minimal Agar

M9N medium and agar was made as M9, except 10 X M9N salts was used in place of 10 X M9 salts.

Luria Broth (LB) 1 % Bactotryptone
0.5 % Yeast Extract
1 % NaCl

LB Agar 1 % Bactotryptone
0.5 % Yeast Extract

1 % NaCl
2 % agar

F Agar

1 % Bactotryptone
0.5 % Yeast Extract
1 % NaCl
5 % glycerol
2 % agar

Glycerol was autoclaved separately and added to the medium prior to pouring the agar plates.

LM Agar

1 % Bactotryptone
0.5 % Yeast Extract
10 mM NaCl
10 mM $MgSO_4$
2 % agar

SOB Medium

2 % Bactotryptone
0.5 % Yeast Extract
10 mM NaCl
2.5 mM KCl
10 mM $MgSO_4$
10 mM $MgCl_2$

$MgSO_4$ and $MgCl_2$ were made up as 1 M stock solutions, autoclaved, and added to the medium after it had been autoclaved.

SOC Medium

SOC medium was as SOB, except that the solution was made to 20 mM glucose after autoclaving.

A.4 Agarose Gel Electrophoresis Buffers

Tris-borate (X10)

0.89 M Boric acid
0.89 M Tris-Cl, pH 8.3
20 mM EDTA

Tris-acetate (X 10)

50 mM NaAc
10 mM EDTA
0.4 M Tris-Cl

The pH was adjusted to 7.5 with glacial acetic acid.

Tracking Dye (X 1)

15 % Ficoll

0.25 % Bromophenol Blue
1 % SDS
0.05 % EDTA, pH 8.5

MOPS Running Buffer (X 5) 0.2 M MOPS, pH 7.0
50 mM NaAc
5 mM EDTA

Loading Buffer (X 1) 20 % Ficoll
0.2 % Bromophenol Blue

Alkaline Gel Buffers

Gel Preparation Buffer (X 1) 50 mM NaCl
1 mM EDTA

Electrophoresis Buffer (X 10) 30 mM NaOH
1 mM EDTA

Loading Buffer (X 1) 50 mM NaOH
1 mM EDTA
2.5 % Ficoll
0.025 % Bromocresol green

A.5 Hybridisation Buffers

10 X SSC 1.5 M NaCl
0.15 M Na citrate

10 X SSCP 1.2 M NaCl
0.15 M Na citrate
0.1 M Na_2HPO_4
0.1 M NaH_2PO_4

10 X SET 1.5 M NaCl
0.3 M Tris-Cl, pH 8.0
10 mM EDTA

20 X SSPE 3.6 M NaCl
0.2 M NaH_2PO_4, pH 7.4
20 mM EDTA

50 X Denhardt's Solution 1 % BSA
1 % Ficoll
1 % Polyvinylpyrrolidone

10 X Nick Translation Buffer 0.5 M Tris-Cl, pH 7.8
50 mM $MgCl_2$

A.6 Hybridisation Selection Buffers

Wash Buffer A
10 mM Tris-Cl, pH 7.6
0.15 M NaCl
1 mM EDTA
0.5 % SDS

Wash Buffer B
10 mM Tris-Cl, pH 7.6
0.15 M NaCl
1 mM EDTA

Neutralisation Mix
10 X SSC
0.5 M Tris
1 M HCl

A.7 Enzyme Buffers

10 X *Sma* I Buffer
200 mM KCl
100 mM Tris-Cl, pH 8.0
100 mM $MgCl_2$
10 mM DTT

All other restriction enzymes were used with the buffer supplied by the manufacturer.

10 X Ligase Buffer
250 mM Tris-Cl, pH 7.6
50 mM $MgCl_2$
25 % Polyethylene glycol 8000
5 mM ATP
5 mM DTT

10 X Polynucleotide Kinase Buffer
0.5 M Tris-Cl, pH 7.6
0.1 M $MgCl_2$
50 mM DTT
1 mM spermidine
1 mM EDTA

Alkaline Phosphatase Dilution Buffer
50 mM Tris-Cl, pH 8.0
0.1 mM EDTA

10 X CIP Buffer (alkaline phosphatase reaction buffer)
0.5 M Tris-Cl, pH 9.0
10 mM $MgCl_2$
1 mM $ZnCl_2$
10 mM spermidine

A.8 RNA Extraction Buffers and Plasmid Preparation Buffers

RNA Extraction Buffers

Buffer A
0.15 M Na acetate, pH 5.0
4 % SDS
20 mM Na iodoacetate
100 µg/ml heparin

Oligo dT Binding Buffer
0.5 M NaCl
10 mM Tris-Cl, pH 7.5
1 mM EDTA
0.1 % SDS

Oligo dT Elution Buffer
10 mM Tris-Cl, pH 7.0
1 mM EDTA
0.1 % SDS

Plasmid Preparation Buffers

Plasmid Maxiprep Solution I
50 mM glucose
25 mM Tris-Cl, pH 8.0
10 mM EDTA

Plasmid Maxiprep Solution II
0.2 M NaOH
1 % SDS

5 M KAc (100 ml)
60 ml 5 M K acetate
11.5 ml glacial acetic acid
28.5 ml H_2O

10 X STE
1 M NaCl
100 mM Tris-Cl, pH 7.8
10 mM EDTA

STET
8 % sucrose
0.5 % Triton-X-100
50 mM EDTA
50 mM Tris-Cl, pH 8.0

TFB
10 mM K-MES, pH 6.2
100 mM KCl
45 mM $MnCl_2$
10 mM $CaCl_2$
3 mM hexamine cobalt (III) chloride

K-MES

K-MES is prepared by adjusting the pH of 1 M MES to 6.2 with KOH.

A.9 Miscellaneous

TE Buffer 10 mM Tris-Cl, pH 7.5
1 mM EDTA

DEAE Elution Buffer 20 mM Tris-Cl, pH 7.6
1 M NaCl

References

References

Ames,B.N., Lee,F.D. and Durston,W.E., 1973, Proc. Natl. Acad. Sci. USA 70 782-786, **An Improved Bacterial Test System for the Detection and Classification of Mutagens and Carcinogens**

Bhattacharyya,A. and Lilley,D.M.J., 1989, Nucleic Acids Research 17 (17) 6823-6838, **The Contrasting Structures of Mismatched DNA Sequences Containing Looped-out Bases (Bulges) and Multiple Mismatches (Bubbles)**

Beck,E. and Zink,B., 1981, Gene 16 35-58, **Nucleotide Sequence and Genome Organisation of the Filamentous Bacteriophages f1 and fd**

Berberich,M.A., 1972, Biochem. Byophys. Res. Commun. 47 (6) 1498-1503, **A Glutamate-delendent Phenotype in *E.coli* K12: The Result of Two Mutations**

Brett,M., Chambers,G.K., Holder,A.A., Fincham,J.R.S. and Wootton,J.C., 1976, J. Mol. Biol. 106 1-22, **Mutational Amino Acid Replacements in *Neurospora crassa* NADP-specific Glutamate Dehydrogenase**

Burns,P.A., Kinnaird, J.H., Kilbey,B.J. and Fincham, J.R.S., 1986, Genetics 113 45-51, **Sequencing Studies of ICR-170 Mutagenic Specificty in the *am* gene of *Neurospora crassa***

Calos,M.P. and Miller,J.H., 1981, J. Mol. Biol. 153 39-66, **Genetic and Sequence Analysis of Frameshift Mutation**

Coddington,M.R., Fincham,J.R.S. and Sundarum,T.K., 1966, J. Mol. Biol. 17 503-512, **Multiple Active Varieties of *Neurospora* Glutamate Dehydrogenase Formed by Hybridisation Between Two Inactive Mutant Proteins *in vivo* and *in vitro***

Cronan Jr.,J.E., Narasimhan,M.L. and Rawlings,M., 1988, Gene 70 161-169, **Insertional Restoration of β-galactosidase α-complementation (White-to-Blue Colony Screening) Facilitates Assembly of Synthetic Genes**

Culbertson,M.R., Charnas,L., Johnson,M.T. and Fink,G.R., 1977, Genetics 86

745-764, **Frameshift and Frameshift Suppressors in** *Saccharomyces cerevisiae*

De Boer,J.G. and Ripley,L.S., 1988, Genetics 118 181-191, **An** *in vitro* **Assay for Frameshift Mutations: Hotspots for Deletions of 1 bp by Klenow- fragment Polymerase Share a Consensus DNA Sequence**

DeMarini,D.M., Pham,H.N., Katz,A.J. and Brockmann,H.E., 1984, Mutation Research 136 185-199, **Relationships Between Structures and Mutagenic Potencies of 16 Heterocyclic Nitrogen Mustards (ICR Compounds) in** *Salmonella typhimurium*

Donahue,T.F., Farabaugh,P.J. and Fink,G.R., 1981, Science 212 455-457, **Suppressible Four-base Glycine and Proline Codons in Yeast**

Ernst,J.F., Hampsey,D.M. and Sherman,F., 1985, Genetics 111 233-241, **DNA Sequences of Frameshift and Other Mutations Induced by ICR 170 in Yeast**

Farabaugh,P.J., Schmeissner,U., Hofer,M. and Miller, J.H., 1978, J. Mol. Biol 126 847-863, **Genetics Studies of the** *lac* **Repressor. VII. On the Molecular Basis of Spontaneous Hotspots in the** *lacI* **Gene of** *Escherichia coli*

Fincham,J.R.S. and Stadler,D.R., 1965, Genet. Res. 6 121-129, **Complementation Relationships of** *Neurospora am* **Mutants in Relation to Their Formation of Abnormal Varieties of Glutamate Dehydrogenase**

Fresco,J.R. and Alberts,B.M., 1960, Proc. Natl. Acad. Sci. USA 46 311-321, **The Accommodation of Non-complementary Bases in Helical Polyribonucleotides and Deoxyribonucleic Acids.**

Frohman,M.A., Dush,M.K. and Martin,G.R., 1988, Proc. Natl. Acad. Sci. USA 85 8998-9002, **Rapid Production of Full-length cDNAs from Rare Transcripts: Amplification Using a Single Gene-specific Oligonucleotide Primer**

Fuchs,R.P.P., Freund,A.M. and Bichara,M., 1988, in **Methods and Consequences of DNA Damage Processing**, eds. Friedberg,E.C. and Hanawalt,P.C., Alan R. Liss, Inc., New York

Fuchs,R.P.P., Levfevre,J.F., Pouyet,J. and Daune,M.P., 1976, Biochemistry 15 3347-3351, **Comparative Orientation of the Fluorene Residue in Native DNA Modified by N-acetoxy-N-2-acetylaminofluorene and Two 7-halogeno Derivatives**

Gama,R.E., Hughes,P.J., Bruce,C.B. and Stanway,G., 1988, Nucleic Acids Research 16 (19) 9346, **Polymerase Chain Reaction Amplification of Rhinovirus Nucleic Acid from Clinical Material**

Gergen,J.P., Stern,R.H. and Wensink,P.C., 1979, Nucleic Acids Research 7 (8) 2115-2136, **Filter Replicas and Permanent Collections of Recombinant DNA Plasmids**

Glickman and Ripley, 1984, Proc. Natl. Acad. Sci. USA 81 512-516 **Structural Intermediates of Deletion Mutagenesis: A Role for Palindromic DNA**

Gough,J.A. and Murray,N.E., 1983, J. Mol. Biol. 166 1-19, Sequence Diversity Among Related Genes for Recognition of Specific Targets in DNA Molecules.

Gubler,U. and Hoffman,E.J., 1983, Gene 25 263-269, **A Simple and Very Efficient Method for Generating cDNA Libraries**

Greener,A., 1990, Stratagene Strategies 3 (1) 5-6, *E.coli* **SURE Strain: Clone "Unclonable" DNA**

Gurr,S.J., Hawkins,A.R., Drainas,C. and Kinghorn,J.R., 1986, Current Genetics 10 761-766, **Isolation and Identification of the** *Aspergillus nidulans gdhA* **Gene Encoding NADP-linked Glutamate Dehydrogenase**

Hanahan,D., 1983, J. Mol. Biol. 166 557-580, **Studies on Transformation of** *Escherichia coli* **with Plasmids**

Hanahan,D. and Meselson,M, 1980, Gene 10 63-67, **Plasmid Screening at High Colony Density**

Hoffmann,G.R., Freemer, C.S. and Parente,L.A., 1989, Mol. Gen. Genet. 218 377-383, **Induction of Genetic Duplications and Frameshift Mutations in** *Salmonella typhimurium* **by Acridines and Acridine mustards: Dependence on Covalent Binding of the Mutagen to DNA**

Holder,A.A., Wootton,J.C., Baron,A.J., Chambers,G.K. and Fincham,J.R.S., 1975, Biochem.J. 149 757-773, **The Amino Acid Sequence of *Neurospora* NADP-specific Glutamate Dehydrogenase. Peptic and Chymotryptic Peptides and Complete Sequence**

Hsieh, C-H. and Griffith,J.D., 1989, Proc. Natl. Acad. Sci. USA 86 4833-4837, **Deletions of Bases in One Strand of Duplex DNA, in Contrast to Single-base Mismatches, Produce Highly Kinked Molecules: Possible Relevance to the Folding of Single-stranded Nucleic Acids**

Ikoku,A.S. and Hearst,J.E., 1981, J. Mol. Biol. 151 245-259, **Identification of a Structural Hairpin in the Filamentous Chimeric Phage M13G*ori1***

Kaguni,L.S. and Clayton,D.A., 1982, Proc. Natl. Acad. Sci. USA 79 983-987, **Template-directed Pausing in *in vitro* DNA Synthesis by DNA Polymerase α from *Drosophila melanogaster***

Kawasaki, E.S., Clark,S.S., Coyne,M.Y., Smith,S.D., Champlin,R., Witte,O.N. and McCormick,F.P., 1988, Proc. Natl. Acad. Sci. 85 5698-5702, **Diagnosis of Chronic Myeloid and Acute Lymphocytic Leukaemias by Detection of Leukaemia-specific mRNA Sequences Amplified *in vitro***

King,P.V. and Blakely,R.W., 1986, BRL Focus 8 (1) 1-3, **Optimizing DNA Ligations for Transformation**

Kinnaird,J.H. and Fincham,J.R.S., 1983, Gene 26 253-260, **The Complete Nucleotide Sequence of the *Neurospora crassa am* (NADP-specific Glutamate Dehydrogenase) Gene**

Kinnaird,J.H., Keighren,M.A., Kinsey,J.A., Eaton,M. and Fincham,J.R.S., 1982, Gene 20 387-396, **Cloning of the *am* (Glutamate Dehydrogenase) Gene *Neurospora crassa* Through the Use of a Synthetic DNA Probe**

Kinsey,J.A., 1977, J. Bacteriology 132 (3) 751-756, **Direct Selective Procedure for Isolating *Neurospora* Mutants Defective in Nicotinamide Adenine Dinucleotide Phosphate-specific Glutamate Dehydrogenase**

Kinsey,J.A. and Fincham, J.R.S., 1979, Genetics 93 577-586, **An Unstable Allele of the *am* Gene of *Neurospora crassa***

Kinsey,J.A., Fincham,J.R.S., Siddig,M.A.M. and Keighren,M., 1980, Genetics 95 249-259, **New Mutational Variants of *Neurospora* NADP-specific Glutamate Dehydrogenase**

Kinsey,J.A. and Hung,B-S.T., 1981, Genetics99 405-414, **Mutation at the *am* Locus of *Neurospora crassa***

Koehl,P., Burnouf,D. and Fuchs,R.P.P., 1989, J. Mol. Biol. 207 355-364 **Construction of Plasmids Containing a Unique Acetylaminofluorene Adduct Located Within a Mutation Hotspot**

Koffel-Schwartz,N. and Fuchs,R.P.P., 1989, Mol. Gen. Genet. 215 306-311, **Genetic Control of AAF-induced Mutagenesis at Alternating GC Sequences: An additional role for RecA**

Koffel-Schwartz,N., Maenhaut-Michel,G. and Fuchs,R.P.P., 1987, J. Mol. Biol. 193 651-659, **Specific Strand Loss in N-2-acetylaminofluorene Modified DNA**

Koffel-Schwartz,N., Verdier,J.M., Bichara,M., Freund,A.M., Daune,M.P. and Fuchs,R.P.P., 1984, J. Mol. Biol. 177 33-51, **Carcinogen Induced Mutation Spectrum in wild type, *uvrA* and *umuC* Strains in *Escherichia coli*: Strain Specificity and Mutation Prone Sequences**

Lederberg,S., 1966, J. Bacteriology 91 (3) 1029-1036, **Genetics of Host-controlled Restriction and Modification of Deoxyribonucleic Acid in *Escherichia coli***

Lee,C.H. and Tinoco,Jr.,I., 1978, Nature 274 609-610, **Mutagen-oligonucleotide Complexes with a Bulged Base as Models for Frameshift Mutation**

Lerman,L.S., 1963, Proc. Natl.Acad. Sci. USA 49 94-102, **The Structure of the DNA-acridine Complex**

Lilley,D.M.J., 1981, Nature 292 380-382, ***In vivo* Consequences of Plasmid Topology**

McPherson,M.J. and Wootton,J.C., 1983, Nucleic Acids Research 11 (15) 5257-5266, **Complete Nucleotide Sequence of the *Escherichia coli gdhA* Gene**

Maniatis,T., Fritch,E.F. and Sambrook,J., 1982, in **Molecular Cloning: A Laboratory Manual**, Cold Spring Harbor Press. Cold Spring Harbor NY.

Mathison,L. and Culbertson,M.R., 1985, Mol. Cell. Biol. 5 2247-2256, **Suppressible and Nonsuppressible +1 G:C Base Pair Insertions Induced by ICR 170 at the *his4* Locus in *Saccharomyces cerevisiae***

Mattaj,I.W., McPherson,M.J. and Wootton,J.C., 1982, FEBS Lett. 147 21-25, **Localisation of a Strongly Conserved Section of Coding Sequence in Glutamic Dehydrogenase Genes**

Mead,D.A., Szczesna-Skorupa,E. and Kemper,B., 1986, Protein Engineering 1 (1) 67-74, **Single-Stranded DNA 'Blue' T7 Promoter Plasmids: A Versatile Tandem Promoter System for Cloning and Protein Engineering**

Moye,W.S., Amuro,N., Rao,J.K.M. and Zalkin,H., 1985, J. Biol. Chem. 260 (14) 8502-8508, **Nucleotide Sequence of Yeast *GDH1* Encoding Nicotinamide Adenine Dinucleotide Phosphate-dependent Glutamate Dehydrogenase**

Nagasu,T. and Hall,B.D., 1985, Gene 37 247-253, **Nucleotide Sequence of the GDH Gene Coding for the NADP-specific Glutamate Dehydrogenase of *Saccharomyces cerevisiae***

Okada,Y. Streisinger,G, Owen,J.(E.),Newton,J., Tsugita,A. and Inouye,M., 1972, Nature 236 338-341, **Molecular Basis of a Mutational Hotspot in the Lysosyme Gene of Bacteriophage T4.**

Owen,J.E., Schultz,D.W., Taylor,A. and Smith,G.R., 1983, J. Mol. Biol. 165 229-248, **Nucleotide Sequence of the Lysosyme Gene of Bacteriophage T4 : Analysis of Mutations Involving Repeated Sequences.**

Pahel,G., Zelenetz,A.D. and Tyler,B.M., 1978, J. Bacteriology 133 (1) 139-148, ***gltB* Gene and Regulation of Nitrogen Metabolism by Glutamine Synthase in *Escherichia coli***

Panayotatos,N. and Wells,R.D., 1981, Nature 289 466-470, **Cruciform Structures in Supercoiled DNA**

Papanicolaou,C. and Ripley,L.S., 1989, J. Mol. Biol. 207 335-353, **Polymerase-specific Differences in the DNA Intermediates of Frameshift Mutagenesis**

Parnes,J.R., Velan,B., Felsenfeld,A., Ramanathan,L., Ferrini,U., Appella,E. and Seidman,J.G., 1981, Proc. Natl. Acad. Sci. USA 78 (4) 2253-2257, **Mouse β_2-Microglobulin cDNA Clones : A Screening Procedure for cDNA Clones Corresponding to Rare mRNAs.**

Porter,A.G., Barber,C., Carey,N.H., Hallewell,R.A., Threlfall,G. and Emtage,J.S., 1979, Nature 282 471-477, **Complete Nucleotide Sequence of an Influenza Virus Haemagglutinin Gene from Cloned DNA**

Pribnow,D., Sigurdson,D.C., Gold, L., Singer,B.S. Napoli,C., Brosius,J., Dull,T.J. and Noller,H.F., 1981, J. Mol. Biol. 149 337-376, ***rII* Cistrons of Bacteriophage T4 : DNA Sequence around the Intercistronic Divide and Positions of Genetic Landmarks.**

Rambosek,J.A. and Kinsey,J.A., 1983, Genetics 105 293-307, **Fine Structure Mapping of the *am* (GDH) Locus of *Neurospora***

Rambosek, J.A. and Kinsey,J.A., 1984, Gene 27 101-107, **An Unstable Mutant Gene of the *am* Locus of *Neurospora* Results from a Small Duplication**

Ricciardi,R.P., Miller,J.S. and Roberts,B.E., 1979, Proc. Natl. Acad. Sci. USA 76(10) 4927-4931, **Purification and Mapping of Specific mRNAs by Hybridisation-selection and Cell-free Translation.**

Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P., 1977, J. Mol. Biol. 113 237, **Labelling Deoxyribonucleic Acid to High Specific Activity *in vitro* by Nick Translation with DNA polymerase I**

Ripley,L.S., 1982, Proc. Natl. Acad. Sci. USA 79 4128-4132, **Model for the Participation of Quasipalindromic DNA Sequences in Frameshift Mutation.**

Ripley,L.S., Glickman,B.W. and Shoemaker,N.B., 1983, Mol. Gen. Genet. 189 113-117, **Mutator Versus Antimutator Activity of a T4 DNA Polymerase Mutant Distinguishes Two Different Frameshifting Mechanisms.**

Ripley,L.S. and Shoemaker,N.B., 1983, Genetics 103 353-366, **A Major Role for Bacteriophage T4 DNA Polymerase in Frameshift Mutagenesis**

Roth,J.R., 1974, Annu. Rev. Genet. 8 319-346, **Frameshift Mutations.**

Sage,E. and Leng,M. 1980, Proc. Natl. Acad. Sci. USA 77 4597-4601, **Conformation of Poly(dG-dC).Poly(dG-dC) Modified by the Carcinogens N-acetoxy-N-2-acetylaminofluorene and N-hydroxy-N-2-aminofluorene**

Sage,E. and Leng,M., 1981, Nucleic Acids Research 9 1241-1249, **Conformational Changes of Poly(dG-dC).Poly(dG-dC) Modified by the Carcinogens N-acetoxy-N-2-acetylaminofluorene**

Saiki,R.K., Scharf,S., Faloona,F., Mullis,K.B., Horn,G.T., Ehrlich,H.A. and Arnheim,N., 1988, Science 230 1350-1352, **Enzymatic Amplification of β-Globin Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anaemia**

Sain,B. and Murray,N.E., 1980, Molec. Gen. Genet. 180 35-41, **The *hsd* (Host Specificity) Genes of *E.coli* K12**

Sakore,T.D., Jain,S.C., Tsai,C.C. and Sobell,H.M., 1977, Proc. Natl. Acad. Sci. USA 74 188-192, **Mutagen-nucleic Acid Intercalative Binding : Structure of a 9-aminoacridine-5-iodocytidyl(3'5')guanosine.**

Sakore,T.D., Reddy,B.S. and Sobell,H.M., 1979, J. Mol. Biol. 135 763-785, **Visualisation of Drug-nucleic Acid Interactions at Atomic Resoluton. IV. Structure of an Aminoacridine-dinucleoside Monophosphate Crystalline Complex, 9-aminoacridine-5-iodocytidyl(3'5')guanosine.**

Sanchez-Pescador,R., Sancivente,E., Valle,F. and Bolivar,F., 1982, Gene 17 1-8, **Recombinant Plasmids Carrying the Glutamate Dehydrogenase Structural Gene from *Escherichia coli* K12**

Sherman,L. and Gefter,M.L., 1976, J.Mol.Biol. 103 61-76, **Studies on the Mechanism of Enzymatic DNA Elongation by *Escherichia coli* DNA Polymerase II**

Short,J.M., Fernandez,J.M., Sorge, J.A., and Huse,W.D., 1988, Nucleic acidsResearch 16 7583-7600, **λ ZAP: A Bacteriophage λ Expression Vector with *in vivo* Excision Properties**

Siddig,M.A.M., Kinsey,J.A., Fincham,J.R.S. and Keighren,M., 1980, J. Mol. Biol. 137 125-135, **Frameshift Mutations Affecting the N-terminal Sequence of *Neurospora* NADP-specific Glutamate Dehydrogenase**

Southern,E., 1975, J. Mol. Biol. 98 503-517, **Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis.**

Stanley,K.K. and Luzio,J.P., 1984, EMBO J. 3 (6) 1429-1434, **Construction of a New Family of High Efficiency Bacterial Expression Vectors: Identification of cDNA Clones Coding for Human Liver Proteins**

Stanolowski Jr.,L.F., Tindall,K.R. and Hsie,A.W., 1986, Mutation Research 160 133-147, **Quantitative and Molecular Analyses of Ethyl Methanesulfonate- and ICR 191-induced Mutation in AS52 Cells**

Streisinger,G. and Owen,J., 1985, Genetics 109 633-659, **Mechanisms of Spontaneous and Induced Frameshift Mutation in Bacteriophage T4**

Streisinger,G., Okada,Y., Emrich,J., Newton,J., Tsugita,A., Terzaghi,E. and Inouye,M., 1966, Cold Spring Harbor Symp. Quant. Biol. 31 77-84,**Frameshift Mutations and the Genetic Code**

Thomas,P.S., 1980, Proc. Natl. Acad. Sci. USA 77(9) 5201-5205, **Hybridisation of Denatured RNA and Small DNA Fragments Transferred to Nitrocellulose.**

Thomas,P.S., 1983, Methods in Enzymology 100 255-266, **Hybridisation of Denatured RNA Transferred or Dotted to Nitrocellulose Paper.**

Valle,F., Becerril,B. Chen,E., Seeburg,P., Heyneker,H. and Bolivar,F, 1984, Gene

27 193-199, **Complete Nucleotide Sequence of the Glutamate Dehydrogenase Gene from** *Escherichia coli* **K-12**

Vieira, J. and Messing,J., 1982, Gene 19 259-268, **The pUC Plasmids, An M13-derived System for Insertion Mutagenesis and Sequencing with Synthetic Universal Primers**

Yanisch-Peron,C., Vieira,J. and Messing,J., 1985, Gene 33 103-119, **Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and pUC 19 Vectors**

Young,R.A. and Davis,R.W., 1983, Proc. Natl. Acad. Sci. USA 80 1194-1198, **Efficient Isolation of Genes Using Antibody Probes**

Additional References

Colgan,D.J., 1988, Nucleic Acids Research 16 10360, **Cloning of Homologous Sequences After Filter Hybridisation**

Love,J. and Deininger,P., 1991, BioTechniques 11 88-92, **Hybridisation Selections, Using Immobilised Driver DNA to Enrich for Clones Unique to a Library**

Lovett,M., Kere,J. and Hinton,L.M., 1991, Proc. Natl. Acad. Sci. USA 88 9628-9632, **Direct Selection: A Method for the Isolation of cDNAs Encoded by Large Genomic Regions**

Parimoo,S., Pantanjali,S.R., Shukla,H., Chaplin,D.D. and Weissman,S.M., 1991, Proc. Natl. Acad. Sci. USA 88 9623-9627, **cDNA Selection: Efficient PCR Approach for the Selection of cDNAs Encoded in Large Chromosomal DNA Fragments**

Pruitt,S.C., 1988, Gene 66 121-134, **Expression Vectors Permitting cDNA Cloning and Enrichment for Specific Sequences by Hybridisation/Selection**