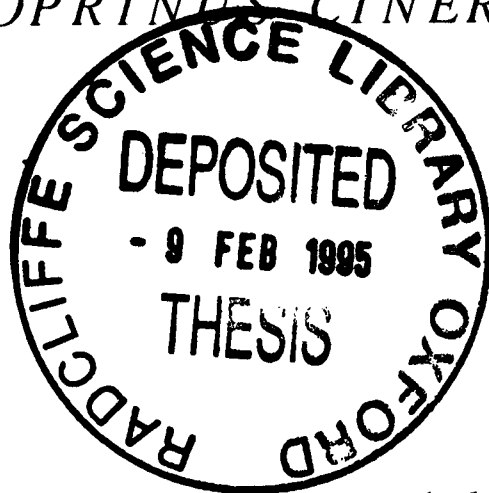


UNIVERSITY OF OXFORD
DEPARTMENT OF PLANT SCIENCES
LINACRE COLLEGE

ANALYSIS OF MATING TYPE
PROTEIN INTERACTIONS
IN *COPRINUS CINEREUS*



A Thesis presented by
Berthold Göttgens
for the degree of Doctor of Philosophy.

April 1994

Acknowledgements

The work described in this thesis was initiated by Dr L.A. Casselton and carried out in her laboratory. Her support and guidance during the project, help with writing it up and the great scientific freedom enjoyed during the time of my D. Phil. are gratefully acknowledged.

I also want to extend my thanks to my second supervisor Dr. J. Mellor for good technical advice and proof-reading of my thesis.

My thanks also go to the other members of Dr Casselton's and Dr Mellor's laboratories for practical and moral support during the time of my D. Phil. and to John Baker for preparing negatives of autoradiographs.

I also want to express my thanks for the financial support by the SERC and the Gathsby Charitable Foundation.

Very special thanks go to my family and Effie for their constant love and support.

to my parents

Analysis of Mating Type Protein Interactions in *Coprinus cinereus*

Berthold Göttgens, Linacre College

D. Phil thesis, Hilary Term 1994

Abstract

The A mating type factor of the hymenomycete fungus *Coprinus cinereus* is a multi-allelic gene complex that controls mating compatibility and sexual development. It contains up to four pairs of specificity genes, the *a*, *b*, *c*, and *d* gene-pairs. Each gene-pair codes for two homeodomain transcription factors with distinct classes of homeodomain motifs. Mating compatibility between the A42 and A6 factors depends solely on the different alleles of the *b* gene-pair, *b1-1* and *b2-1* in A42 and *b1-3* and *b2-3* in A6. The *b1-3* and *b2-3* genes of A6 were isolated and the complete DNA sequences of genomic and cDNA clones were determined. Construction of chimeric genes using the A42 and A6 *b* genes identified the N-terminal regions of the A proteins as being responsible for allele specificity. Analysis of protein-protein interactions showed that b1 and b2 proteins from different alleles of the same gene-pair can dimerise, whereas proteins from the same allele pair can not. It was shown that a region of 90 amino acids at the N-terminus of the b2-3 protein is sufficient for dimerisation with b1-1. This region is predicted to contain an amphipathic helix. A comparison with the equivalent region in the b2-1 protein identifies a similar helix. This suggests that a compatible A mating type reaction and thus allele specificity is recognised by the ability to dimerise through this domain. Polyclonal antibodies were raised against the b1-1 protein and a heterologous yeast expression system was established for testing potential DNA target sites of the b1-1 and b2-3 proteins, both techniques offering potentially useful tools for further molecular analysis of the A mating type proteins.

Chapter 1	1
Introduction	1
1.1 Breeding systems of the fungi.....	1
1.2 Life cycle of <i>Coprinus cinereus</i>	3
1.3 Structure of the A and B mating type factors	5
1.4 Regulation of Sexual Development by A and B.	6
1.6 The molecular structure of the A mating type factor of <i>C. cinereus</i>	7
1.7 Molecular structure of the A mating type factor	9
1.7.1 Homeodomain proteins and their role in determining fungal mating types.....	11
1.8 General aims of the thesis.....	16
Chapter 2	18
Materials and Methods	18
2.1 Strains, plasmids and vectors.	18
2.1.1 <i>C. cinereus</i> strains.....	18
2.1.2 Bacterial Strains.	18
2.1.3 <i>Saccharomyces cerevisiae</i> strain.....	19
2.1.4 Cloning Vectors.	20
2.1.5 DNA sequences from the A42 factor of <i>C. cinereus</i>	21
2.2 Chemicals and Stock solutions.	21
2.2.1 Chemicals.....	21
2.2.2 Stock Solutions.....	22
2.3 Media.....	22
2.3.1 <i>C. cinereus</i> Media and Media stock solutions	23
2.3.2 Bacterial media.....	24
2.3.3 Yeast media.....	25
2.3.4 Antibiotics.....	26
2.4 Culturing Conditions.	26
2.4.1 <i>Coprinus cinereus</i> culturing conditions.....	26

2.4.2	Bacterial culturing conditions.....	27
2.4.3	<i>S. cerevisiae</i> culturing conditions.....	27
2.5	DNA-mediated transformation.....	28
2.5.1	DNA-mediated transformation of bacteria.....	28
2.5.2	Transformation of <i>C. cinereus</i> (modified from (Casselton and de la Fuente Herce 1989)).....	29
2.5.3	Transformation of <i>Saccharomyces cerevisiae</i>	30
2.6	DNA isolation and manipulation.....	32
2.6.1	Plasmid DNA isolation.....	32
2.6.2	Isolation of plasmid DNA for transformation of <i>S.</i> <i>cerevisiae</i>	32
2.6.3	Isolation of single stranded DNA.....	32
2.6.4	Estimation of RNA and double-stranded DNA concentration.....	33
2.6.5	Restriction enzyme digestion of DNA	33
2.6.6	Agarose gel electrophoresis of DNA.....	34
2.6.7	Photography of agarose gels.....	34
2.6.8	Isolation of DNA fragments from agarose gels.....	35
2.6.9	Cloning Procedures.....	35
2.7	Southern blot analysis	36
2.7.1	Southern blotting (Southern 1975).....	36
2.7.2	Hybridisation procedure.....	36
2.7.4	Washing of filters	37
2.7.5	Autoradiography.....	38
2.7.6	Re-probing of Hybond-N filters.	38
2.8	DNA-sequencing.....	38
2.8.1	DNA-sequencing protocol	38
2.8.2	DNA sequence analysis.....	39
2.9	RNA isolation and manipulation.....	40

2.9.1	RNA Extraction from <i>Coprinus cinereus</i> using the Guanidinium Thiocyanate method.....	40
2.9.2	Isolation of poly (A)+ RNA.....	41
2.9.3	RNA agarose gel electrophoresis.....	43
2.10	Northern blot analysis.....	44
2.10.1	Northern blotting.....	44
2.10.2	Northern hybridisation.....	45
2.10.3	Washing of filters.....	45
2.10.4	Stripping and re-probing of Northern filters.....	46
2.11	Construction of a cDNA library	46
2.11.1	Synthesis of cDNA	46
2.11.2	Cloning of cDNA.....	48
2.11.3	Maintenance of the cDNA library.....	49
2.11.4	Screening of the cDNA library	50
2.12	Screening of the LambdaGEM 11 genomic library	51
2.12.1	Plating and screening the library.....	51
2.12.2	Small scale isolation of λ GEM 11 phage DNA	51
2.12.3	Phage DNA purification.	52
2.13	Gel electrophoresis of proteins	53
2.13.1	SDS-PAGE minigels.....	53
2.13.2	Preparative SDS-PAGE.....	54
2.14	Expression of genes in pGEX vectors (Smith and Johnson 1988).....	55
2.14.1	Small scale expression	55
2.14.2	Large scale isolation of HDX3 GST fusion protein	56
2.14.3	Large scale inclusion body preparation.....	56
2.14.4	Large scale isolation of soluble GST protein.....	57
2.14.5	Expression in the presence of sorbitol and betaine.....	57
2.14.6	Thrombin cleavage of fusion proteins	58
2.15	Expression of genes in the pT7-7 vector.....	59

2.15.1	Expression using a two plasmid system.....	59
2.15.2	Expression using <i>E. coli</i> BL21(DE3) strains	59
2.17	Preparation of antisera	60
2.17	Western blot analysis	60
2.18	<i>In vitro</i> transcription/translation	61
2.19	Crosslinking of <i>in vitro</i> translated proteins.....	61
2.20	Immunoprecipitation of <i>in vitro</i> translated protein.....	62
2.21	Preparation of <i>C. cinereus</i> crude protein extracts.....	63
2.23	Preclearing of antisera.....	63
2.23	Immunofluorescence with fungal mycelia	64
2.24	GST fusion protein in vitro association assay	65
Chapter 3		66
Cloning of the <i>b</i> gene pair of the A6 mating type factor		66
3.1	Screening a genomic library for genes of the A6 mating type factor.....	67
3.2	Subcloning of the A6 mating type locus genomic DNA fragments in pBluescript II, KS-.....	71
3.3	Identification of the A6 <i>b</i> genes by transformation.....	75
3.4	Preparation of a cDNA library of <i>C. cinereus</i> strain H9	76
3.5	Screening of the H9 cDNA library for clones of the A6 <i>b</i> genes	79
Chapter 4		82
4.1	Sequencing of the A6 <i>b1-3</i> and <i>b2-3</i> genes.....	82
4.2	Gene structure as revealed by primary sequence data	87
4.2.1	Start sites	95
4.2.2	Introns.....	96
4.2.3	Promotor sequences.....	97
4.2.4	Polyadenylation.....	98
4.3	Analysis of the predicted protein sequences of the <i>b1-3</i> and <i>b2-3</i> genes	98

4.3.1 The b1-3 protein.....	98
4.3.2 The b2-3 protein.....	102
4.4 Detailed analysis of the homeodomains	105
4.5 Construction of chimeric b genes of the <i>C. cinereus</i> A42 and A6 mating type factors.....	111
4.6 Mapping the region between the <i>b1-3</i> and the <i>d1-1</i> genes of the A6 mating type factor.....	116
Chapter 5	119
Raising polyclonal antibodies against the <i>C. cinereus</i> b1-1 mating type protein.....	119
5.1 Plasmid constructs for overexpression of the A42 <i>b</i> gene- pair in <i>E. coli</i> using T7 RNA polymerase systems.....	119
5.1.1 Constructs for expression of <i>b1-1</i> using the pT7-7 vector system	119
5.1.1.1 Expression studies with <i>b1-1</i> cDNA in pT7- 7.....	121
5.1.2. Expression studies with the 5' end of <i>b2-1</i> cloned into pT7-7.....	125
5.2 Plasmid constructs for overexpression of the A42 <i>b</i> genes in <i>E. coli</i> using the pGEX vector system	127
5.2.1 Preparation of a construct containing the 5' half of the <i>b1-1</i> gene	127
5.2.2 New plasmid constructs for <i>b1-1</i> starting at the second ATG	132
5.3 Plasmid construct for overexpressing <i>b2-1</i> gene sequence using the pGEX vector system.....	136
5.4 Immunisation of two rabbits with the XB1 fusion protein.....	139
5.5 Analysis of antisera	140
5.5.1 Analysis of antisera by Western blotting.....	140

5.5.1.1 Analysis of antisera by western blotting using the XB1 fusion protein.....	140
5.5.1.2 Analysis of antisera by Western blotting using <i>C. cinereus</i> total protein extracts.....	141
5.5.2 Analysis of antisera by immunoprecipitation of in vitro translated b1-1 protein	143
5.5.3 Analysis of antisera by immunofluorescence.....	145
Chapter 6	146
Analysis of protein-protein interactions of the <i>C. cinereus</i> A mating type proteins	146
6.1 Analysis of protein-protein interactions by chemical crosslinking.....	149
6.1.1 Construction of a full length <i>b2-1</i> cDNA clone	150
6.1.2 Construction of a full length <i>b2-3</i> cDNA clone for <i>in</i> <i>vitro</i> translation	151
6.1.3 Chemical crosslinking studies with the b1-1, b2-1 and b2-3 proteins	152
6.2 Analysis of protein-protein interactions by co- immunoprecipitation.....	154
6.3 Detection of protein-protein interactions using b1-1-GST fusion proteins and in vitro translated <i>C. cinereus</i> A mating type proteins	156
6.3.1 Initial XB1 b2-3 association experiment.....	158
6.3.2 Identification of potential dimerisation regions of the b1-1 and the b2-3 proteins using the GST association assay.....	159
6.3.3 Assay for interactions between the XB1 and <i>in vitro</i> translated b1-1 protein using the GST association technique.....	163

6.3.4 Analysis of XB1/b2-1 protein-protein interactions using the GST <i>in vitro</i> association assay.....	165
Chapter 7	167
Expression of the <i>C. cinereus</i> A mating type proteins in <i>S. cerevisiae</i>	167
7.1 Construction of plasmids	170
7.2 Transformation of <i>S. cerevisiae</i> with the <i>C. cinereus</i> mating type genes.....	174
Chapter 8	178
Discussion.....	178
8.1 Gene-pair organisation.....	178
8.2 Conserved features of b1 and b2 proteins	179
8.3 The N-terminus of the b proteins determines allele specificity	182
8.4 The N-terminus of b1 and b2 proteins contains a dimerisation domain	183
8.5 Analysis of the N-terminal domains of b1 and b2 proteins	186
8.6 Mechanism of a compatible A mating type reaction	196
References	202
Appendix I	221
Appendix II	230

Abbreviations

<i>ade</i>	adenine
ATP	adenosine triphosphate
bHLH	basic helix loop helix
bHLHZIP	basic helix loop helix zipper
bp	base pair
cDNA	copy deoxy ribonucleic acid
cpm	counts per minute
Δ	deletion
dATP	deoxy adenosine triphosphate
dCTP	deoxy cytosine triphosphate
dGTP	deoxy guanosine triphosphate
DMSI	dimethyl suberimidate
DNA	deoxy ribonucleic acid
dNTP	deoxynucleotide triphosphate
dTTP	deoxy thymidine triphosphate
ECL	enhanced chemiluminescence
frac.	fraction
g	gram
h	hour
HD1	<i>C. cinereus</i> homeodomain 1
HD2	<i>C. cinereus</i> homeodomain 2
HLH	helix loop helix
IgG	immunoglobulin G
kbp	kilo base pair
kd	kilo dalton
l	litre
M	molar

mg	milli gram
ml	milli litre
mM	milli molar
mm	milli meter
mRNA	messenger ribonucleic acid
nm	nano meter
o/n	over night
OD	optical density
<i>pab</i>	para amino benzoic acid
PAGE	polyacrylamide gel electrophoresis
pfu	plaque forming units
polyA	poly adenosine
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
s.s.	single stranded
TCA	trichloro acetic acid
TM	trademark
trp	tryptophan
UV	ultra violet
vol	volume
w/v	weight to volume
μg	micro gram
μl	micro litre
μm	micro meter

Chapter 1

Introduction

1.1 Breeding systems of the fungi

The majority of organisms possess genetic barriers to reduce inbreeding and to promote outbreeding. The unicellular ascomycete *Saccharomyces cerevisiae* possesses one mating type locus with two different allelic forms, MAT_a and MAT_α. This divides the population into two cross-compatible haploid cell types, a-cells and α-cells. Of the four haploid cells derived from meiosis of a diploid a/α cell, two are a and two are α, thereby limiting sib-compatibility to 50%. In basidiomycete fungi, more complex genetic systems have evolved in which overall restrictions on sib-compatibility can be reduced to 25% but outbreeding potential can exceed 90%. This is because the mating type genes of the basidiomycetes can be multi-allelic. In the case of the hymenomycete *Coprinus cinereus*, the subject of this thesis, it is estimated that there are some 12,000 different mating types (Whitehouse 1949) (Raper 1966). Outbreeding is thought to be genetically advantageous because it maximises genetic recombination in the population.

Classical genetic studies (Whitehouse 1949) (Raper 1966) established that there were two incompatibility systems in the hymenomycetes. Some 25% of species studied had a bipolar (unifactorial) system in which mating type is determined by a single locus. 65% had a tetrapolar (bifactorial) system with mating type determined by two unlinked loci. The remaining 10% were found to be self compatible but assumed to be secondarily homothallic (Whitehouse 1949).

In bipolar species, the single mating type locus was designated A. This has multiple specificities i.e. A1, A2, A3 etc. all of which are cross-compatible. From a single fruit body, two mating types segregate in equal numbers (hence the term bipolar). In tetrapolar species, the two unlinked mating type loci were designated A and B . These were also shown to have multiple specificities, i.e. A1, A2, A3, etc., B1, B2, B3, etc. and for compatibility, both A and B must be different. Four mating types segregated with equal frequency from a single fruit body (hence the term tetrapolar).

The efficiency of a breeding system is measured in terms of inbreeding restriction and promotion of outbreeding. The advantage of multiple mating types in this context is shown in table 1. Mating compatibility in the hymenomycetes is only recognised intracellularly after the hyphal fusion, and it becomes very important to restrict inbreeding and to promote outbreeding in order to reduce the chances of non-productive (incompatible) hyphal fusions.

Table 1.1: Efficiency of bipolar and tetrapolar breeding systems in restricting inbreeding and promoting outbreeding (After (Casselton 1978))

Incompatibility System	Genes	Meiotic Segregation	Inbreeding Restriction (%)	Outbreeding Potential Alleles @ each locus		
				A	B	(%)
Bipolar	$A_1 A_2.....A_n$	0.5 A_1	50	2		50
		0.5 A_2		5		80
				10		90
				20		95
Tetrapolar	$A_1 A_2.....A_n$ $B_1B_2.....B_n$	0.25 A_1B_1	25	2	2	25
		0.25 A_2B_2		5	5	64
		0.25 A_1B_2		10	10	81
		0.25 A_2B_1		20	20	90

The table shows that a $A_1B_1 \times A_2B_2$ dikaryon produces spores of four different compatibility groups or mating types. These germinate into monokaryons with inbreeding restriction reduced to 25% and an outbreeding potential of 90%, provided that 20 alleles for each mating type factor are present in the total population (species).

The tetrapolar system is more efficient than the bipolar one in reducing sib-incompatibility to 25% and even with two loci involved, it requires very few different alleles of *A* and *B* to approach almost 100% outbreeding potential. In fact it requires only 20 alleles of both *A* and *B* to give 90% outbreeding potential. The advantage of such high outbreeding potential is to impose little restriction on general genetic recombination. It is however still surprising to discover that there may be as many as 160 different *A* and 79 *B* specificities in *C. cinereus* (Whitehouse calculation used by (Raper 1966)). In another hymenomycete, *Schizophyllum commune*, it is estimated that there are 288 different *A* and 81 different *B* specificities (Raper 1966).

1.2 Life cycle of *Coprinus cinereus*.

The life cycle of *Coprinus cinereus* is illustrated in Figure 1.1. The products of successful sexual reproduction, the basidiospores, are borne on gills produced on the undersurface of the fungal fruit body. Each spore germinates to establish a sexually sterile primary mycelium, termed the monokaryon, which is capable of continuous growth by apical extension (Casselton 1978). The monokaryon has uninucleate cells and produces abundant uninucleate asexual spores, known as oidia. Sexual development is initiated when two compatible monokaryons fuse to produce the secondary mycelium, termed the dikaryon.

The dikaryon has binucleate cells, one nucleus derived from each monokaryon, and characteristic clamp connections at each septum.

During the vegetative growth of the dikaryon, the synthetic and morphogenetic activities of each cell of the mycelium are under the joint control of the two associated nuclei. Like the monokaryon, the dikaryon is capable of continuous vegetative growth and is very stable in nature. It no longer produces oidia but differentiates the fruit bodies.

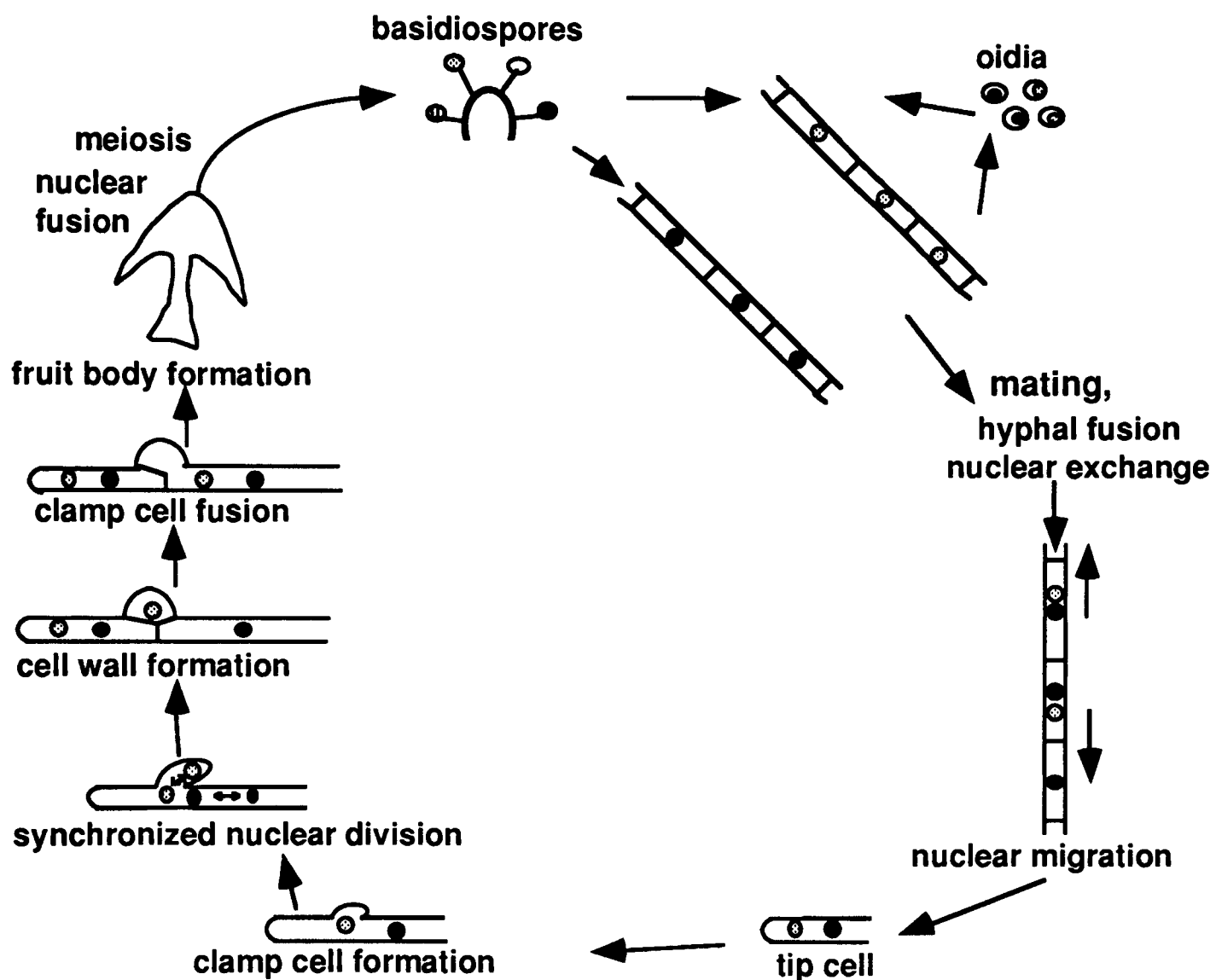


Figure 1.1: Life cycle of *Coprinus cinereus* (detailed description in text).

Several steps can be distinguished in the development of the dikaryon following cell fusion. Nuclei are exchanged and in each case the donor nucleus migrates through the established cells of the recipient monokaryon (See Figure 1.1). Nuclear migration necessitates the breakdown of the complex dolipore septa between each cell (Giesy and Day 1965), and continues until the hyphal tip cells have two genetically different nuclei. A complex cell division occurs in all dikaryotic cells, during which a short lateral bump forms on the side of the apical cell, which elongates and bends away from the cell tip. This specialised cell is called a clamp cell. One

nucleus enters the clamp cell and the two nuclei then undergo a synchronised division. New septa are laid down thereby generating three distinct cells, a binucleate apical cell, a uninucleate subapical cell and a small uninucleate clamp cell. The clamp cell fuses with the subapical cell and its nucleus migrates into this cell to re-establish the dikaryotic nuclear pair. The formation of the clamp cell ensures equal distribution of the two genetically different nuclei to each daughter. Given the correct light and temperature regime, mushroom fruit bodies develop on the secondary mycelium. The dikaryotic pairs of nuclei fuse to form a transient diploid phase in the basidia or spore-bearing cells derived from the fertile hymenial tissue on the gills of the mushroom. Meiosis follows immediately after nuclear fusion to produce four haploid nuclei which migrate singly into the four basidiospores formed on the outer surface of the basidium.

1.3 Structure of the *A* and *B* mating type factors

Classical recombination analysis established that the *A* and *B* mating types of *S. commune* (Papazian 1950) (Raper, Baxter et al. 1958) (Raper, Baxter et al. 1960) and at least the *A* mating type of *C. cinereus* (Day 1960) were determined by two closely linked genes termed α and β . This led to the use of the term 'factor' rather than gene to describe both *A* and *B* (Raper 1966). The *A α* and *A β* loci of *S. commune* are approximately 8 map units apart and recombination between the two loci is easily detected (Koltin, Stamberg et al. 1972) (Giasson, Specht et al. 1989). Both α and β were shown to be multiallelic and for *A* factors to be different they needed to differ in alleles at only one locus. Thus, $\alpha1\beta1$, $\alpha2\beta2$, $\alpha1\beta2$ and $\alpha2\beta1$ all have different *A* specificities. Since α and β are functionally redundant it was suggested (Raper 1966) that they may have arisen by gene duplication. Recombination studies with isolates of *S. commune* collected from all over the world

identified 9 $A\alpha$ and 32 $A\beta$ alleles. Similarly for the B factor, the α and β loci were found to be 0.1 to 8 map units apart and 9 $B\alpha$ and 9 $B\beta$ alleles were identified. Recombination analysis in *C. cinereus* (Day 1960) confirmed the existence of $A\alpha$ and $A\beta$ loci and showed that they were 0.07 map units apart. A similar analysis of the B factor, however, failed to show a two gene structure (Haylock, Economou et al. 1980).

1.4 Regulation of Sexual Development by A and B

A compatible mating requires A and B factors to be different but there are no barriers to fusions between cells with common factors. Forced matings between partially compatible monokaryons having the same A or B factor showed that each factor independently regulates different parts of the morphogenetic sequence that gives rise to the dikaryon (Swiezynski and Day 1960). An incompatible fusion can give rise to a heterokaryon in which cells contain two different nuclei but not in a regular dikaryotic association. There are three possible incompatible associations: common A , common B and common AB . The common A and common B heterokaryons have been useful tools for determining the regulatory roles of the A and B mating type genes in both *C. cinereus* and *S. commune* (Raper 1966) (Casselton 1978) and in both fungi A and B appear to have equivalent regulatory functions.

The B factor controls the nuclear migration which precedes the establishment of the dikaryon. This involves the disruption of the complex dolipore septa, to allow the free passage of the donor nuclei through the recipient mycelium (Swiezinsky and Day 1960) (Giesy and Day 1965) (Raper 1966). The A mating type factor regulates the development of clamp cells and synchronised nuclear division in the binucleate tip cells (Swiezynski and Day 1960) (Raper, Baxter et al. 1960) (Raper 1966). The fusion of the clamp cell to the subapical cell is a B regulated function. In a mating with

different *A* factors but common *B* factors, unfused or false clamps are formed where the clamp cell fails to fuse with the sub-apical cell and its nucleus is trapped leaving the subapical cell uninucleate.

The work described in this thesis is concerned with the structure and function of the *A* mating type genes of *C. cinereus*. The development of false clamp cells was the phenotype used throughout to assay a functional cloned *A* mating type gene (see below). In a common *A* mating where there are different *B* mating type factors nuclear exchange and nuclear migration can occur but no clamp cells are formed. When both *A* and *B* factors are common, as expected no morphological changes can be detected (Raper 1966) (Swiezinsky and Day 1960) (Swiezynski and Day 1960). Heterokaryons are generally infertile.

1.6 The molecular structure of the *A* mating type factor of *C. cinereus*

A question which could not be resolved by classical genetic analysis was how such large numbers of mating types are generated in the hymenomycetes. Recent molecular studies of the *A* mating type locus of *C. cinereus* has provided some answers to this question but now raises new questions as to how the products of the mating type genes regulate mating compatibility and sexual development.

The first cloned *A* mating type factor of *C. cinereus* was isolated in 1989 (Mutasa, Tymon et al. 1990). The *A* factor is flanked by two closely linked metabolic genes pab-1 and ade-8 (Day 1960) as shown in figure 1.2. This provided the strategy for isolating *A* from a cosmid genomic library by a chromosome walk from the metabolic marker pab-1 and made it possible to distinguish between the functionally redundant α and β genes.



Figure 1.2: Linkage of the A mating type factor of *C. cinereus* to the metabolic genes pab1 (p-aminobenzoic acid synthesis) and ade8 (adenine biosynthesis). Distances between loci are given in map units (based on (Day 1960)).

Cloned sequences containing functional A42 genes led to the formation of unfused clamp cells after DNA mediated transformation into a suitable *C. cinereus* host strain (mating type A5 or A6). The unfused clamp cell phenotype is shown in figure 1.3. As transforming DNA generally integrates non-homologously in *C. cinereus*, the transformants were able to express both the resident A factor (A5 or A6) and the ectopically integrated A42 gene(s).

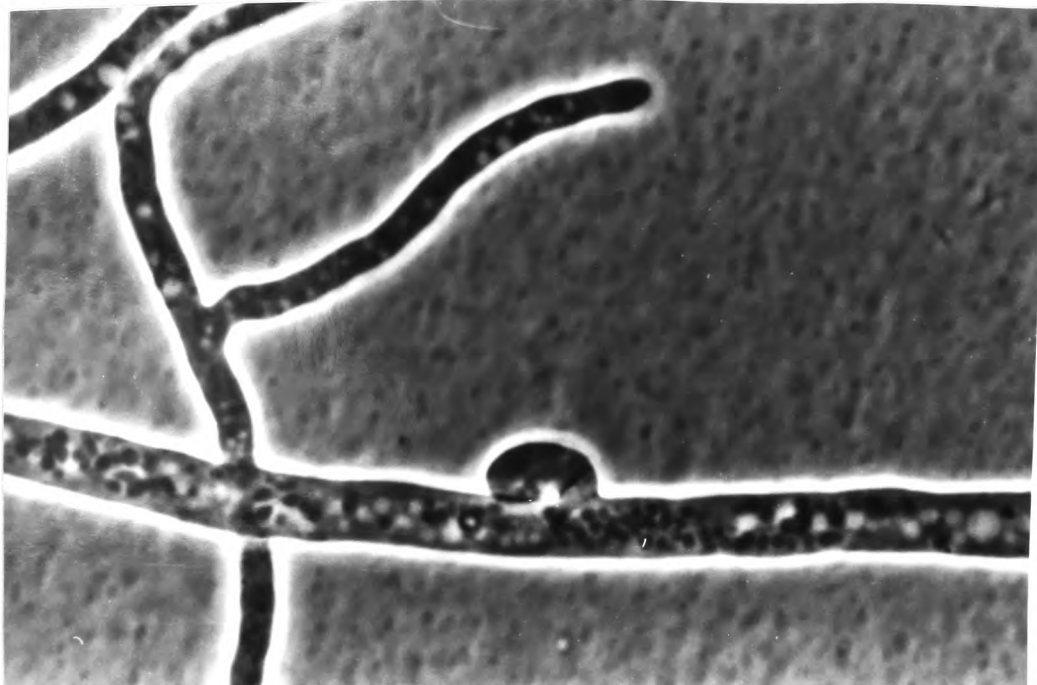


Figure 1.3: Development of unfused clamp cells by an A6 monokaryon following transformation with a cosmid clone containing the A42 mating type factor.

1.7 Molecular structure of the A mating type factor

The molecular analysis of the A42 factor revealed an organisation much more complex than anticipated from the classical genetic analysis (Kües, Richardson et al. 1992). Seven genes were identified separated into two sub-complexes that correspond to the classical $A\alpha$ and $A\beta$ loci (see figure 1.4). The two flanking genes, one in the α locus (α flanking gene (α -fg)) and the other in the β locus. (β flanking gene (β -fg)), appear from hybridisation data to have similar DNA sequence in all A factors studied. Preliminary DNA sequence analysis of the α -flanking gene identified a motif typical for metalloendoproteinases. The function of both flanking genes and whether they play a role in mating is not clear. The other genes have multiple alleles with DNA sequences so dissimilar that they do not cross-hybridise in Southern analysis to each other or to different allelic forms present in other A factors. These genes determine the different specificities of A and have been called the specificity genes. The α and β loci are separated by a non-coding homologous DNA sequence which has been called the homologous hole. This sequence is 7 kbp in length and corresponds to the 0.07 map units that was shown by classical analysis to separate $A\alpha$ and $A\beta$ (Day 1960).

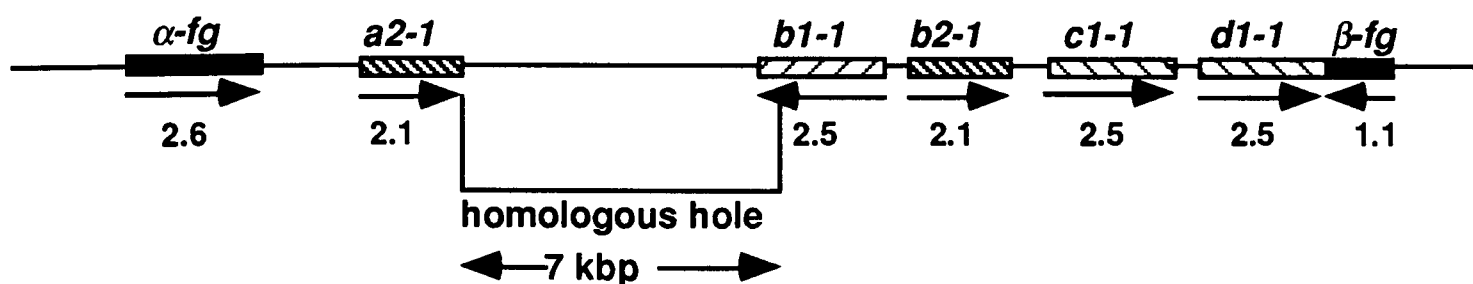


Figure 1.4: Molecular organisation of the A42 mating type factor of *C. cinereus* showing orientation and size (in kb) of the transcripts of the flanking and specificity genes (Kües, Richardson et al. 1992) (for nomenclature of the specificity genes see below)

Two classes of specificity genes were shown to be present in the A42 factor. These were initially recognised by transcript sizes of 2.1 kb and 2.5 kb. DNA sequence analysis revealed that the two different message sizes corresponded to two classes of proteins with distinctly different homeodomains. The longer genes encode the homeodomain motif designated HD1 and the smaller genes a so called HD2 homeodomain (Kües, Richardson et al. 1992). The homeodomain is a 60 amino acids long domain that characterises a class of eukaryotic transcription factors (Scott, Tamkum et al. 1989) and it mediates DNA binding to their target sites (Pabo and Sauer 1992). Structural analysis of several homeodomains and homeodomain-DNA complexes demonstrated a conserved secondary structure with an extended N-terminal arm and three α -helices (Quian, Billeter et al. 1989) (Otting, Qin et al. 1990) (Kissinger, Liu et al. 1990) (Wolberger, Vershon et al. 1991). Helix 1 and helix 2 pack against each other in an anti-parallel arrangement. Helix 3 is roughly perpendicular to the first two helices and fits into the major groove of the recognition sequence. The main contacts to the DNA target site are provided by helix 3 and the extended N-terminal arm. Helix 1 and helix 2 only provide some contacts to the DNA backbone (see figure 1.5).

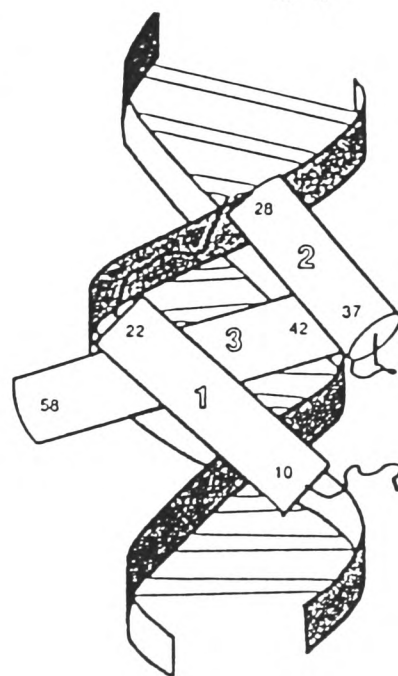


Figure 1.5: Sketch of the engrailed homeodomain-DNA complex (Kissinger, Liu et al. 1990), summarising the overall relationship of the three α -helices and the N-terminal arm with respect to the DNA recognition sequence.

1.7.1 Homeodomain proteins and their role in determining fungal mating types

The first evidence for the involvement of homeodomain proteins in the regulation of sexual compatibility came from sequence analysis of the MAT locus of *Saccharomyces cerevisiae*. Two distinct classes of homeodomain motifs were found in the a_1 and the α_2 mating type proteins. *S. cerevisiae* has two different mating types called **a** and α (Herskowitz 1989); **a**-cells contain the a_1 and a_2 protein and α -cells the α_1 and α_2 proteins respectively (Kües and Casselton 1992). These proteins have been shown to act as transcription factors (Nasmyth and Shore 1987) (Herschbach and Johnson 1993), regulating mating type specific gene expression in haploid yeast cells. Homodimers of the α_2 protein in conjunction with the general transcription factor MCM1 repress expression of a-specific genes in α -haploid cells (Smith and Johnson 1992). After mating of an a- and an α -cell, the a_1 and α_2 proteins heterodimerise to form a new transcription factor activity, which controls the phenotypic switch from the haploid to the diploid stage (Dranginis 1990).

The hemibasidiomycete *Ustilago maydis* has two mating type loci, called *a* and *b*. The *a* locus encodes pheromones, while DNA sequence analysis of the *b* mating type locus revealed the same two classes of homeodomain proteins that were found in the *C. cinereus* A mating type factor (Schulz, Banuett et al. 1990) (Gillissen, Bergemann et al. 1992). The *U. maydis* *b* mating type locus is multi-allelic and its organisation is much simpler than the *C. cinereus* A factor. It contains just one divergently transcribed pair of genes, one of which codes for an HD1 (called *bE* gene) and the other one for an HD2 (called *bW* gene) protein. Different strains are compatible if they have different alleles of this gene-pair at the *b* mating type locus (Gillissen, Bergemann et al. 1992). DNA sequence analysis of allelic forms

of the *b* genes showed that the predicted proteins only differed in no more than the 120 N-terminal amino acids. This results in a protein structure with a variable N-terminal domain followed by the conserved homeodomain and a constant C-terminal domain. It was further established that compatibility was determined by an interaction of an HD1 protein of one gene-pair with an HD2 protein of a different allelic gene-pair as shown in figure 1.6 (Gillissen, Bergemann et al. 1992). As for *S. cerevisiae*, again it is the interaction of two classes of homeodomain proteins that leads to the switch from unmated to mated phenotype. However, each *b* mating type locus of *U. maydis* already contains the two classes of genes, and evolution has ensured that HD1 genes do not interact with the HD2 gene belonging to the same pair.

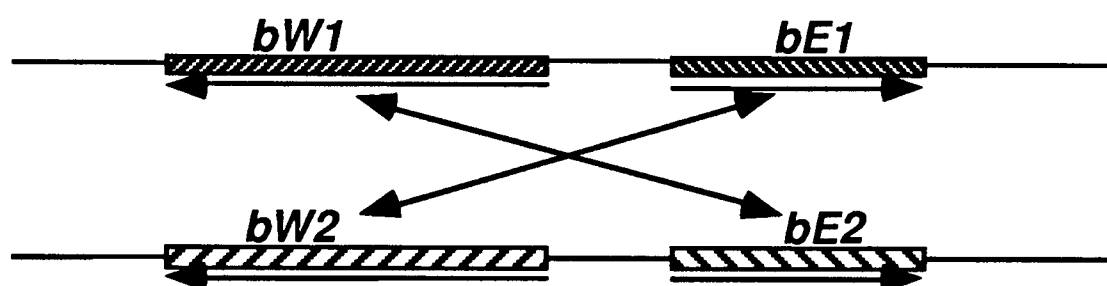


Figure 1.6: Interaction of HD1 and HD2 genes at the *b* mating type locus of *U. maydis*. The figure shows the organisation and the interaction of the *b1* and *b2* alleles

Molecular analysis of the $A\alpha$ mating type factor of the basidiomycete *S. commune* showed a similar picture to the *U. maydis* *b* mating type locus, i.e. one pair of divergently transcribed genes coding for homeodomain proteins of the HD1 and HD2 class (Stankis, Specht et al. 1992). The HD1 proteins were called Z proteins and the HD2 proteins Y proteins. This analysis showed that in the basidiomycetes one gene of a gene-pair may be missing, i.e. the $A\alpha1$ locus of *S. commune* contains only an HD2 gene coding for the Y1 protein. DNA mediated transformation of Y and Z genes into the $A\alpha1$ background showed that , as for *U. maydis*, the two different

classes of genes had to interact in a compatible mating (Specht, Stankis et al. 1992).

These findings in other fungi and the comparison of several cloned *A* factors suggested a model for an archetypal *A* factor in *C. cinereus* (Kües and Casselton 1992) (Kües and Casselton 1993). It contains four pairs of divergently transcribed HD1 and HD2 specificity genes as illustrated in figure 1.7. The $A\alpha$ locus contains the α flanking gene and one pair of specificity genes. The $A\beta$ locus contains three pairs of specificity genes and the β flanking gene. The four pairs of specificity genes have been designated the *a*, *b*, *c* and *d* gene-pair, with HD1 genes called *a1* to *d1* and HD2 genes *a2* to *d2*.

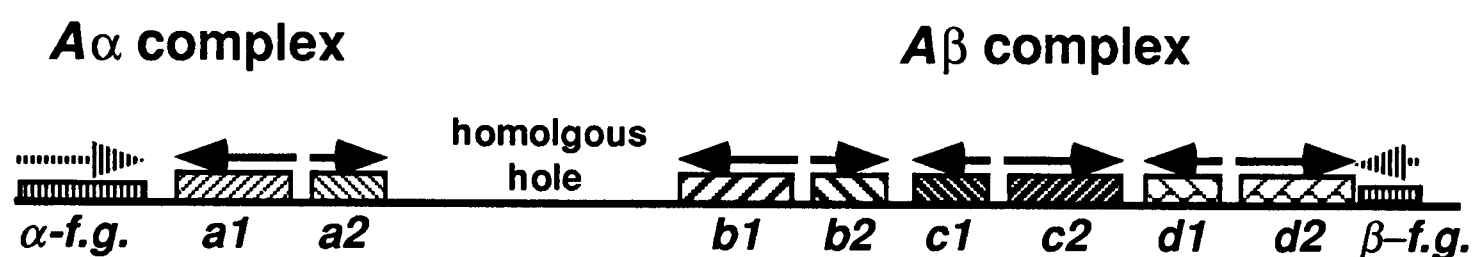


Figure 1.7: Model of the *C. cinereus* archetypal *A* mating type factor

This complex structure was assumed to have evolved from an *U. maydis* like structure with only one pair of specificity genes by duplication and diversification of pairs of specificity genes (Kües and Casselton 1993).

The analysis of different $A\alpha$ factors gave evidence that compatibility in mating was similar to the situation in *U. maydis*: The genes of a particular gene-pair had to be from different alleles to allow successful mating. It was further postulated that proteins of one gene-pair could not interact with proteins of a different gene-pair in order to bring about a compatible mating reaction. This means that *a* genes only positively interact with a different allele of an *a* gene and not with a *b*, *c* or *d* gene. All *A* factors that were

subsequently analysed showed that some of the genes postulated for the archetype were missing (Kües and Casselton 1993) maybe reflecting the low selection pressure to maintain all genes if they are functionally redundant. This result is in line with previously mentioned data for the $A\alpha 1$ locus of *S. commune* which only contains one mating type gene. In summary, the molecular analysis of the *C. cinereus* A factor showed striking similarities to the *b* mating type locus of *U. maydis* but with several more gene-pairs. The presence of several gene-pairs has the advantage that few alleles are needed to create many different mating type specificities. For example *C. cinereus* could generate 27 different mating types with three alleles each of three different gene-pairs. That is a total of 9 alleles of the specificity gene-pairs which would only yield nine different mating types in *U. maydis*.

Evidence for compatible interactions being confined to alleles of the same gene-pair is provided by classical analysis showing that the $A\alpha$ and $A\beta$ locus can be recombined through crossing over at the homologous hole sequence. This suggests, that any genes from the $A\alpha$ locus can be recombined with any genes at the $A\beta$ locus which would inevitably lead to self compatible strains if $A\alpha$ genes were to interact with $A\beta$ genes in mating. However, recombination experiments never yielded self compatible strains (Day 1963). Taking advantage of the fact that only functionally identical alleles cross-hybridise in Southern blot analysis hybridisation studies with genomic DNA of *C. cinereus* strains of various mating types has shown, that within the $A\beta$ locus, alleles of the *b*, *c* and *d* gene-pairs can exist in numerous combinations (Kües, Richardson et al. 1992). This would again lead to self compatibility if genes from different gene-pairs could interact to bring about a successful mating.

Additional evidence for the interaction of HD1 and HD2 proteins came from an independent analysis. Rare mutations conferring self-compatibility

map within the A locus (Day 1960). Such mutations are particularly interesting because they show constitutive expression of the A regulated development of clamp cells. In these mutant cells, A regulated development no longer requires compatible HD1 and HD2 proteins to be brought together by mating. Cloning (Richardson 1992) and DNA sequencing of a self compatible A factor (A6 mut) showed that it contains a single gene derived from a fusion of an HD2 (*a2-1*) and an HD1 (*d1-1*) specificity gene (Kües and Casselton 1992). This mutant gene confers self compatibility to any *C. cinereus* strain when introduced by transformation and it has been proposed that the covalent fusion of an HD2 and an HD1 protein in the mutant protein suggests, that in *C. cinereus*, as in *U. maydis* and *S. commune*, the phenotypic switch from the monokaryon to the dikaryon is brought about through the interaction of HD1 and HD2 proteins (Kües and Casselton 1992).

At the start of this study only the A42 (Kües, Richardson et al. 1992), the A43 (May, Le Chevanton et al. 1991) and parts of the A6 mating type factors (Richardson 1992) of *C. cinereus* had been cloned.

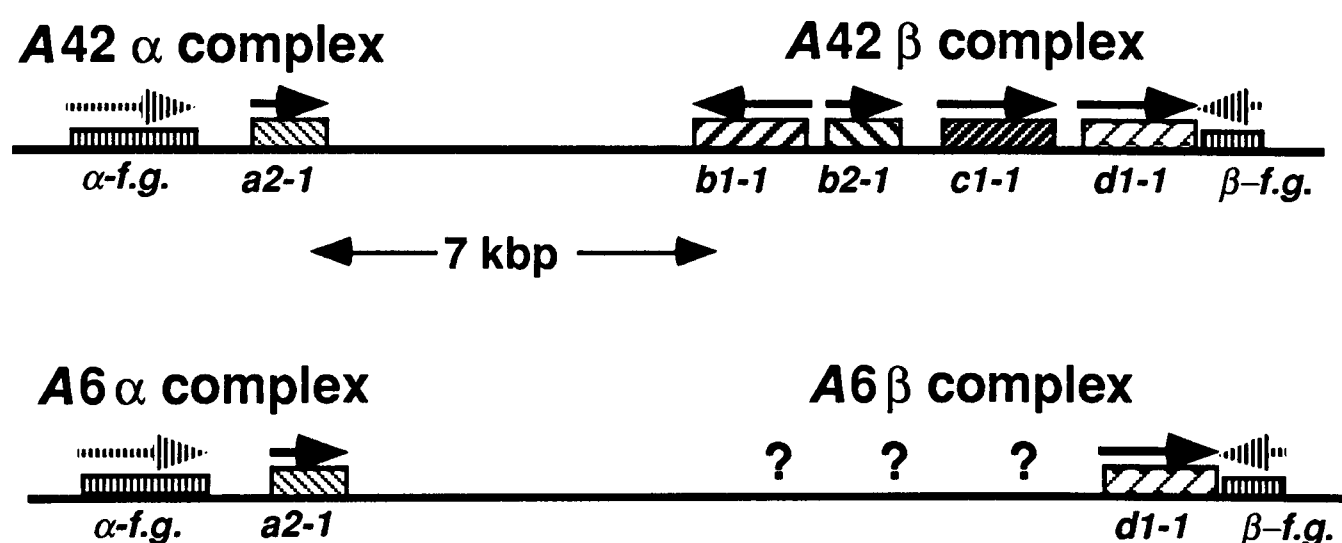


Figure 1.8: Comparison of the *C. cinereus* A42 and A6 mating type factors as known at the beginning of this study

Figure 1.8 gives a comparison of the A42 mating type factor and fragments of A6. Based on the proposed archetype (figure 1.6), A42 has one HD2 gene from the *a* gene-pair at the *Aα* locus (*a2-1*), a complete *b* gene-pair (*b1-1* and *b2-1*), one HD1 gene from the *c* gene-pair (*c1-1*) and one HD1 gene from the *d* gene-pair (*d1-1*). It is obvious that both factors share the entire *Aα* locus and also the *d1-1* gene. Transformation studies of cloned A42 genes into an A6 strain of *C. cinereus* has shown that only the *b1-1* and the *b2-1* gene were able to switch on clamp cell formation in this assay (Kües, Richardson et al. 1992). It was therefore concluded that compatibility of the A42 and A6 mating type factors relies solely on the presence of different alleles of the *b* gene-pair. The *c1-1* gene was inactive in the A6 strain and several other strains tested which suggested that this gene may not be functional or that A6 did not contain any *c* genes that *c1-1* could interact with.

Analysis of the compatible interaction between the different *b* gene alleles present in the *C. cinereus* A42 and A6 mating types should provide an experimental system of reduced complexity and allow the discovery of some basic mechanisms that allow a compatible mating.

1.8 General aims of the thesis.

The first aim of this study was to isolate and sequence the *b* gene-pair from the A6 mating type factor of *C. cinereus*, and to compare the DNA sequences and predicted protein sequences with those of the *b* gene-pair from A42. Despite sequence dissimilarity one could expect some conserved sequence or structural features suggesting functionally important domains within the proteins.

The second aim was to express cDNAs of the *b* genes of A42 and A6 in *E. coli* and *in vitro* to allow functional studies with the *C. cinereus* A proteins.

The main aim was to analyse interactions of compatible and incompatible proteins, such as dimerisation. This would lead to the development of a powerful system for studying homeodomain protein function in a eukaryotic differentiation process.

Such information is useful for a better understanding of homeodomain protein function in higher eukaryotes. Homeodomain proteins in these organisms have been isolated either because of mutation studies or via DNA sequence similarity and there is still little information as to their mode of action (Appel and Sakonju 1993) (Manak and Scott 1993) (Budd and Jackson 1991).

Chapter 2

Materials and Methods

2.1 Strains, plasmids and vectors.

2.1.1 *C. cinereus* strains.

C. cinereus strains used in this thesis are detailed in Table 2.1.

Table 2.1: *C. cinereus* strains

Strain	Mating Type	Auxotrophic Marker(s)	Origin
H9	<i>A₆B₆</i>	wild-type	Lewis
LT2	<i>A₆B₆</i>	<i>trp-1</i>	Casselton
JV6	<i>A₄₂B₄₂</i>	wild-type	Hedger
NA2	<i>[A]</i> null <i>B₆</i>	<i>ade-8</i> , <i>trp1.1,1.6</i>	Pardo
LN118	<i>A₄₂B₄₂</i>	<i>ade-2</i> , <i>trp-1.1,1.6</i>	Casselton

2.1.2 Bacterial Strains.

All bacterial strains of *Escherichia coli* used in the current study are listed in table 2.2.

Table 2.2: Genotype of Bacterial strains.

STRAIN	GENOTYPE
DH5 α	F ⁻ endA1 hsdR17 (r _k ⁻ m _k ⁺) supE44 thi-1 recA1 gyrA96 relA1 Δ (lacZA-argF) Δ lacU169 (Φ 80 Δ lacZ Δ M15) (Hanahan 1983)
WM1704	Δ lac V169 Δ Ion araD139 chr::Tn10 hfl A150 rpsl
SURE TM	e14 ⁻ (mcrA) Δ (mcrCB-hsdSMR-mrr)171 sbcC recB recJ umuC::Tn5(kan ^r) uvrC supE44 lac gyrA96 relA1 thi-1 endA1 [F' proAB lacI ^q Z Δ M15] (Stratagene)
SOLR TM	e14 ⁻ (mcrA) Δ (mcrCB-hsdSMR-mrr)171 sbcC recB recJ umuC::Tn5(kan ^r) uvrC lac gyrA96 relA1 thi-1 endA1 λ^R [F' proAB lacI ^q Z Δ M15] Su ⁻ (non suppressing) (Stratagene)
BL21(DE3)	F ⁻ ompT hsdSB (rB ⁻ mB ⁻ ; an <i>E. coli</i> B strain) with a λ prophage carrying the T7 RNA polymerase (Studier, Rosenberg et al. 1990)
BL21(DE3) LysS	F ⁻ ompT hsdSB (rB ⁻ mB ⁻ ; an <i>E. coli</i> B strain) with a λ prophage carrying the T7 RNA polymerase + pLysS (Studier, Rosenberg et al. 1990)
BL21(DE3) LysE	F ⁻ ompT hsdSB (rB ⁻ mB ⁻ ; an <i>E. coli</i> B strain) with a λ prophage carrying the T7 RNA polymerase + pLysE (Studier, Rosenberg et al. 1990)
KW251	F ⁻ supE44 supF58 galK2 galT22 metB1 hsdR2 mcrB1 mcrA- argA81:Tn10 recD1014 (Promega).
XL1-Blue	supE44 hsdR17 recA1 endA1 gyrA46 thi relA1 lacF'[proAB ⁺ lacI ^q lacZ Δ M15 Tn10(tet ^r)] (Stratagene)

2.1.3 *Saccharomyces cerevisiae* strain

The *S. cerevisiae* strain YPH500 used in the current study was kindly provided by Alex Andrianopoulos (University of Georgia, Athens Georgia). The genotype is *MAT α* , *ura3-52*, *lys2-801*, *ade2-101*, *trp1- Δ 63*, *his3- Δ 200*, *leu2- Δ 1*.

2.1.4 Cloning Vectors.

All cloning vectors used are listed in table 2.3.

Table 2.3: Cloning vectors used in the current study.

vector	description	reference
pBluescript-II, KS+/-, SK+/-	used for subcloning and sequencing	(Stratagene)
λ GEM11	cloning vector of H9 genomic library	(Promega)
YRP12	cloning vector of H9 genomic library	(Scherer and Davis 1979)
pT7-7	T7 expression vector	(Tabor and Richardson 1985)
pGP1-7	encodes T7 RNA polymerase	(Tabor and Richardson 1985)
pGEX2T	expression vector	(Smith and Johnson 1988)
pGEX3X	expression vector	(Smith and Johnson 1988)
λ ZAP TM	cDNA library construction	(Stratagene)
pAA54	centromeric yeast expression vector (<i>HIS3</i> marker)	(A. Andrianopoulos)
pAA55	centromeric yeast expression vector (<i>TRP1</i> marker)	(A. Andrianopoulos)
pYC7	2 μ m based reporter gene vector	(Chang and Timberlake 1992)
pYC8	2 μ m based reporter gene vector	(Chang and Timberlake 1992)

2.1.5 DNA sequences from the A42 factor of *C. cinereus*

DNA sequences from the A42 mating type factor of *C. cinereus* were used for screening the genomic library derived from the A6B6 strain H9 (chapter 3) and for the construction of plasmids for expression in *E. coli*, *S. cerevisiae* and *in vitro* (chapters 5,6 and 7). Figure 2.1 shows the origin of the A42 sequences used for the preparation of expression constructs.

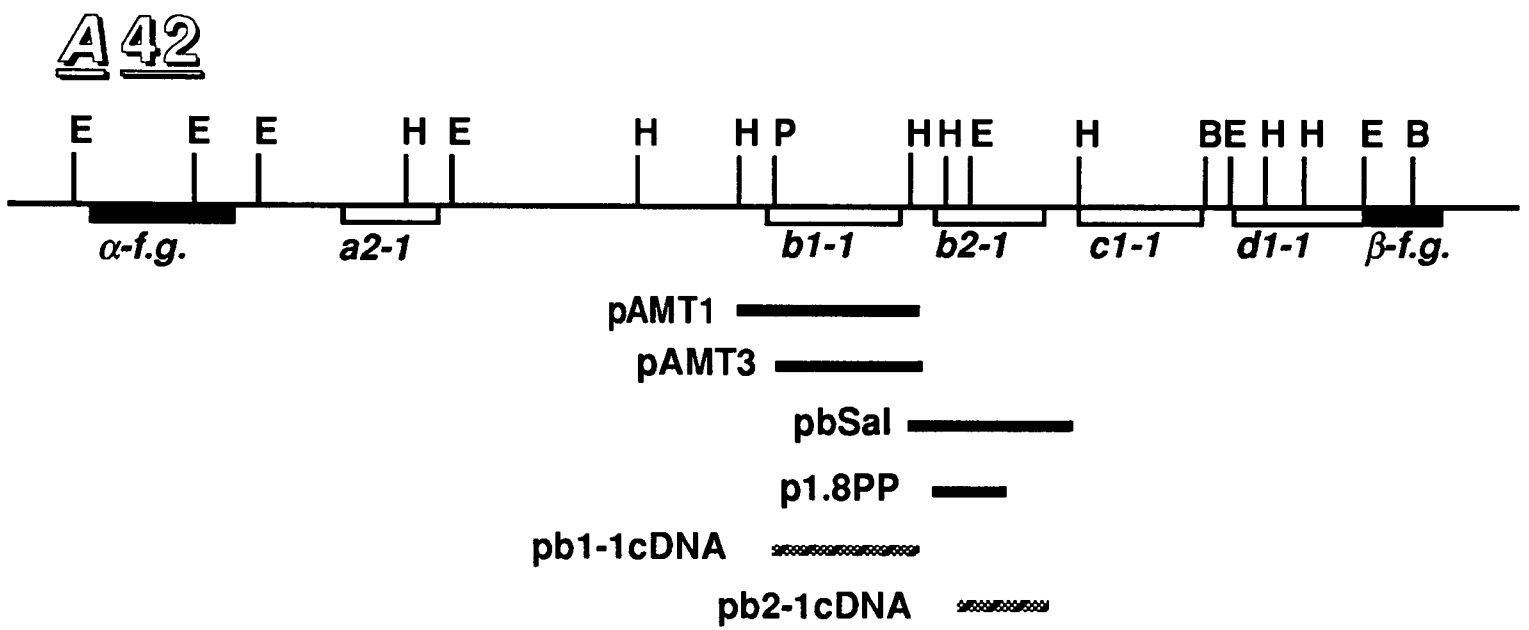


Figure 2.1: Origin of DNA sequences from the A42 mating type factor used for the preparation of expression constructs. Position of fragments shown in relation to a restriction map of the A42 factor (solid lines = genomic DNA fragments, shaded lines = cDNA fragments).

2.2 Chemicals and Stock solutions.

2.2.1 Chemicals.

All chemicals used in the preparation of media and buffers were of analytical grade.

2.2.2 Stock Solutions.

The following stock solutions were prepared according to (Sambrook, Fritsch et al. 1989):

Denhardt's reagent (50x); 0.5 M EDTA (pH8.0); Ethidium Bromide (10mg/ml); deionized Formamide; IPTG(0.1 M); PBS; Phenol/chloroform; Proteinase K (20 mg/ml); RNase A (10 mg/ml); Salmon sperm DNA (sheared and denatured); SDS 10% or 20%); SEVAG; 3 M Sodium acetate (pH4.8 or pH5.2); 0.5 M Sodium Phosphate Buffer (pH 7.0); SM Buffer; SSC (20x); SSPE; 50 x TAE (Tris-acetate electrophoresis buffer); 5 x TBE; TBS; 10 x TE (pH 8.0); 10x TNE; Tris-HCl buffers; 2% X-Gal

Sephadex G-50 was prepared as follows: 30 g Sephadex G-50 was suspended in 250 ml TNE/0.2% SDS, autoclaved, cooled and an equal volume of sterile TNE/0.2% SDS added. Stored at 4°C, warming to room temperature before use.

2.3 Media

Bacterial media were prepared with deionized water, whereas analytical grade water was used for all *C. cinereus* and yeast culture media. Media were sterilised by autoclaving at 121°C for 15 minutes.

2.3.1 *C. cinereus* Media and Media stock solutions

Trace element stock solution

1.0 g Calcium chloride, anhydrous, 4.1 g Magnesium chloride, hydrate, 0.53 g Iron (III) citrate, hydrate, 0.44 g Magnesium sulphate, hydrate, 0.4 g Zinc sulphate, hydrate dissolved in 400 ml water and stored at 4 °C.

Minimal Medium (Shariari and Casselton 1974)

2.0 g L-asparagine, 10.0 g Glucose, 0.5 g Ammonium tartrate, 1.0 g Potassium dihydrogen orthophosphate, 2.25 g Disodium hydrogen orthophosphate, 0.29 g Sodium sulphate, 40.0 µg Thiamine, 0.25 g Magnesium sulphate dissolved in 1L water. 15.0 g of agar added before autoclaving.

***C. cinereus* Complete Medium**

10.0 g Glucose, 2.0 g L-asparagine, 0.75 g Difco-Bacto Yeast extract, 0.75 g Difco-Bacto casitone, 1.13 g Difco-Bacto Malt extract, 0.1 g Adenine sulphate, 0.5 g Ammonium Tartrate, 1.0 g Potassium dihydrogen phosphate, 0.29 g Disodium sulphate, anhydrous, 2.25 g Disodium hydrogen phosphate, anhydrous, 40 µg Thiamine, aneurine HCl, 0.25 g Magnesium sulphate dissolved in 1000 ml water. 15.0 g agar added prior to autoclaving.

Complete Medium (RNT) (Rao and Niederpruem 1969)

4.0 g Difco-Bacto Yeast extract, 10.0 g Difco-Bacto Malt extract, 4.0 g Glucose, 100.0 mg L-Tryptophan dissolved in 1000 ml water. 10.0 g agar added before autoclaving.

Regeneration Medium. (Binninger, Skrzynia et al. 1987)

5.0 g Glucose, 2.0 g L-Asparagine, 1.0 g Potassium dihydrogen orthophosphate, 2.25 g Disodium hydrogen orthophosphate, 40.0 µg Thiamine, 0.25 g Magnesium sulphate, 172.0 g Sucrose, 5.0 g soluble starch dissolved in water and the volume adjusted to 1,000 ml. 15.0 g Agar added before autoclaving.

Liquid Medium.

Liquid minimal medium was prepared as above, with the omission of magnesium sulphate and agar, but with the addition of 40 ml/L *Coprinus* trace element solution. Liquid RNT was as above but without addition of agar.

Supplemented medium

When required, supplements were added to media at the following concentrations per litre of medium: L-Tryptophan (0.1 g/l), Adenine sulphate (0.1 g/l).

2.3.2 Bacterial media.**LB (Luria-Bertani) medium.**

LB medium was prepared according to (Sambrook, Fritsch et al. 1989). For solid LB, 1% agar was added prior to autoclaving. LB glycerol contained 15% glycerol and LB top agarose 0.8% agarose, added prior to autoclaving.

NZY medium

All NZY liquid and solid media were prepared according to the instructions in the λ ZAP cloning kit manual (Stratagene)

X-gal medium.

2.5 g Sodium chloride, 2.5 g Potassium chloride, 10.0 g Difco-Bacto Tryptone, 1.0 ml 1M Calcium chloride in 1000 ml water, add 10.0 g agar. After autoclaving, 100 mg ampicillin (Sigma A-9518, sodium salt), 2 ml of 2% X-gal and 2 ml 0.1 M IPTG was added.

2.3.3 Yeast media**YEPD**

1% yeast extract, 1% peptone, 2% glucose; sterilised by autoclaving. For solid medium, 2% agar was added.

SC (synthetic medium)

0.67% YNB (yeast nitrogen base w/o amino acids(Difco)), 2% glucose, 2% agar; sterilised by autoclaving

SSX medium (colour detection)

For 500 ml of SSX medium, 3.35 g of YNB and 1.4 g of agar were autoclaved in 400 ml AnalaR water. Then 50 ml of sterile 1 M potassium phosphate buffer pH7, 50 ml of sterile 20% galactose, 1 ml of X-Gal stock and 10 ml of amino acids stock solution were added. Plates were poured after adding 5 ml of any additional amino acids.

Amino acids (nutrients) stock solution

Adenine, arginine, tyrosine, aspartic acid and methionine all at 0.1%. Lysine and phenylalanine at 0.15%. Isoleucine at 0.3%, glutamic acid at 0.5%, valine at 0.75%, threonine at 1% and serine at 2%. All were dissolved in AnalaR water and autoclaved. 2 ml of this stock were used for every 100 ml of synthetic medium

Additional amino acids/nutrients

0.2% tryptophan (stored at 4°C), 0.2% histidine, 0.2% leucine and 0.1% uracil were dissolved individually in AnalaR water and autoclaved. 1 ml of the individual stocks was added for every 100 ml of synthetic medium or SSX.

2.3.4 Antibiotics.

Antibiotic stock solutions were prepared and stored as described below:

Ampicillin, 50 mg/ml. Ampicillin (Sigma A-9518) dissolved in sterile water and stored at -20°C.

Chloramphenicol, 34 mg/ml. Chloramphenicol (Sigma C-0378) dissolved in ethanol and stored at -20°C.

Kanamycin, 50 mg/ml. Kanamycin sulphate (Sigma K-4000) dissolved in sterile water stored at -20°C.

Tetracycline, 2 mg/ml. Tetracycline (Sigma T-3383) dissolved in ethanol:water (1:1 v/v) and stored in the dark at -20°C.

2.4 Culturing Conditions.

2.4.1 *Coprinus cinereus* culturing conditions.

Mycelial cultures of *C. cinereus* were grown at 37°C on solid medium in 90 mm plastic Petri dishes. These were stored for limited periods at 4°C. For long term storage, 5 mm x 5 mm blocks of mycelium were placed in 10% glycerol and stored at -70°C. Small-scale liquid cultures were initiated by inoculating 5 mm x 5 mm blocks of mycelium grown on complete or RNT solid medium into 15-20 ml of liquid medium in Petri dishes. These cultures were incubated at 37°C for 5 days. For large-scale liquid cultures

Petridish cultures were macerated for 1 minute in short bursts at half speed in a MSE Atomix blender. The macerate was inoculated into 200 ml liquid medium (with the addition of chloramphenicol at a final concentration of 34 µg/ml) in 1,000 ml conical flasks and incubated at 37°C on an orbital shaker (speed 100-150 rpm).

2.4.2 Bacterial culturing conditions.

To initiate bacterial cultures, cells were streaked onto LB agar medium (containing the necessary antibiotics if plasmids were to be maintained) and incubated at 37°C.

Small-scale liquid cultures were grown by inoculating single colonies into 5 ml liquid medium (plus any necessary antibiotics) in a boiling tube. The culture was incubated on a roller drum overnight at 37°C to reach stationary phase. To obtain log-phase cultures, the bacteria were grown for 2-3 hours (depending on the strain), until an OD₆₀₀ of 0.5–0.7 was reached.

For long term storage of bacteria, cells were pelleted from overnight cultures and resuspended in 1 ml LB containing 15% glycerol and the appropriate antibiotic, and stored at -70°C.

2.4.3 *S. cerevisiae* culturing conditions

To initiate yeast cultures, cells were streaked onto solid medium and incubated inverted at 30°C overnight or until the cultures had grown sufficiently. Strains transformed with prototrophic markers were kept on synthetic medium containing necessary supplements. Individual colonies from these plates were used to inoculate liquid cultures which were grown at 30°C on an orbital shaker (120-150 rpm).

2.5 DNA-mediated transformation.

2.5.1 DNA-mediated transformation of bacteria.

Preparation of competent cells

0.2–0.3 ml of a fresh overnight culture was inoculated into 250 ml of liquid LB medium. This was vigorously shaken at 150–200 rpm at 37°C until the cells reached an OD₅₅₀ of 0.2-0.5.

The cells were chilled in ice water for 10 minutes and harvested by centrifugation at 4000 x g for 10 minutes at 4°C. The cell pellet was resuspended in 125 ml of cold 0.05M calcium chloride and 10 mM Tris-HCl pH 8.0 and incubated on ice for a minimum of 2 hours to induce competence. The cells were harvested at 4000 x g for 10 minutes and resuspended in 20 ml cold 0.05 M calcium chloride. After adding 7 ml cold sterile glycerol, the cells were gently mixed and 0.5 ml aliquots were dispensed into pre-chilled 1.5 ml Eppendorf tubes. The competent cells were stored at -70°C.

Transformation of bacterial cells (Dagert and Ehrlich 1979)

The frozen competent cells were defrosted on ice. An appropriate amount of DNA (0.1-30 ng) was added to 100 µl competent cells and incubated on ice for 10 minutes. Following a brief heat shock at 37°C for 5 minutes, 1 ml of pre-warmed liquid LB was added and the cells incubated at 37°C for one hour. An appropriate amount of cells (10-100 µl) was plated on to selective medium. The plates were incubated overnight at 37°C.

This procedure was shortened for retransformations of plasmid DNA or for transformations into strains carrying the pGP1-7 plasmid. In these cases,

the defrosted cells and the DNA were mixed and incubated on ice for 3 minutes and subsequently plated out on selective media (turbo-transformation).

2.5.2 Transformation of *C. cinereus* (modified from (Casselton and de la Fuente Herce 1989)).

Preparation of protoplasts.

MM buffer 0.5 M Mannitol, 50 mM Sodium maleate (pH5.5).

MMC buffer 0.5 M Mannitol, 50 mM Sodium maleate (pH5.5), 50 mM Calcium chloride.

Digestion buffer 40 mg/ml cellulase (Onozuka R10) and 1 mg/ml chitinase (Sigma C-6137) dissolved in sterile MM buffer and sterilised by centrifugation (17k at 4°C for 30 minutes: SS-34 rotor, Sorvall RC-3C).

To harvest the oidia, 10 ml of sterile water was poured on to the surface of a five day old culture and the surface gently scraped with a sterile blade. Mycelial fragments were removed by filtering through sterile glass wool. Cells were washed in MM buffer, and resuspended in 2-4 ml digestion buffer and incubated on a roller drum at 37°C for 3-5 hours.

Protoplasts were recovered by centrifugation at 2.5 K in a bench-top centrifuge at room temperature for 5 minutes, washed twice in MM and once in MMC and finally resuspended in an appropriate volume of MMC.

DNA-mediated transformation

PEG (polyethylene glycol) PEG 4000 25% (w/v), 10 mM Tris-HCl (pH 7.5), 25 mM Calcium chloride

STC 1 M Sorbitol, 50 mM Calcium chloride, 10 mM Tris-HCl (pH 7.5).

For each transformation, 1–5 µg of DNA (in 10 µl TE) and 12.5 µl PEG were added to 50 µl of protoplast suspension. All transformations carried out in the current study were cotransformations, i.e. the plasmid to be transformed was mixed with 1 µg of the plasmid pCc1001 (Binniger, Skrzynia et al. 1987), which contains the *C. cinereus trp-1* gene, prior to mixing with the protoplasts. The tube contents were mixed gently and incubated on ice for 20 minutes. 0.5 ml PEG was added and the tube contents mixed by inversion. The transformation mix was incubated for 5 minutes at room temperature after which time 1 ml of STC was added followed by gentle mixing by inversion. Cells were plated on to the surface of regeneration agar. Host strains for transformation contained a non-revertible mutation in the *trp-1* (tryptophan synthetase) structural gene that could be complemented by transformation with a cloned *trp-1* gene. Trp⁺ transformants were selected and examined for expression of A-regulated clamp cell development. Transformants appeared after 3-4 days incubation at 37°C.

2.5.3 Transformation of *Saccharomyces cerevisiae*

S. cerevisiae strain YPH500 was transformed using the lithium acetate transformation procedure following the procedure in (Ito, Kukuda et al. 1983) with modifications.

Preparation of competent cells

LiOAc 0.1 M LiOAcetate in 1 x TE prepared by mixing appropriate amounts of sterile 10 x stock solutions and sterile AnalaR water.

Competent yeast cells were prepared by inoculating a single colony into 10 ml of YEPD medium. After 24 hours growth at 30°C, 2×10^6 cells from this culture were used to inoculate a fresh 100 ml YEPD culture. It was assumed that an OD₆₀₀ of 1.0 is equivalent to 1×10^7 cells per ml. After a further 16 hours at 30°C, the cells were pelleted by centrifugation at 3,000 rpm for 5 min in a bench top centrifuge. The cells were then washed with 20 ml of TE buffer and pelleted as above. The cells were resuspended in 0.1 M LiOAc/TE solution at a density of 5×10^8 cells per ml and incubated at 30°C with gentle shaking for 1 hour. 100 µl aliquots of these cells were used for each transformation.

DNA-mediated transformation

PEG/LiOAc 40% PEG 3350 dissolved in 0.1 M LiOAcetate and filter sterilised

100 µg of carrier DNA (10 mg/ml sonicated salmon sperm DNA stock) and 0.1 - 2 µg of the chosen plasmid DNA were added to each 100 µl aliquot of competent cells. 1 - 2 µg of each plasmid were added for cotransformations. The cells were incubated for 30 min at 30°C with gentle shaking. 0.7 ml of PEG/LiOAc solution were added to each transformation and after a further incubation at 30°C for 30 min, the cells were treated for 10 min at 42°C. Cells were then pelleted by centrifugation in a microfuge (2 seconds at full speed) and washed twice with 1 ml TE (pelleting as above). The washed cell pellets were resuspended in 100 µl TE, spread on the appropriate selection

plates and incubated inverted at 30°C until the colonies reached the required size. Colonies were transferred to SSX plates by replica plating and incubated for up to 7 days to monitor any appearance of blue colonies. Plates were stored at 4°C.

2.6 DNA isolation and manipulation.

2.6.1 Plasmid DNA isolation

Plasmid DNA was purified from bacterial cultures following the alkaline lysis protocol as described in (Sambrook, Fritsch et al. 1989). The yield was 25-50 µg plasmid DNA/ 3 ml overnight culture for bluescript based plasmids, 10-25 µg for the *E. coli* expression vectors and the yeast vectors. DNA required for sequencing and transformation of *C. cinereus* was digested with RNase (3 µl of stock solution per miniprep) for 30 min, extracted with phenol/chloroform and ethanol precipitated.

2.6.2 Isolation of plasmid DNA for transformation of *S. cerevisiae*

Plasmid DNA used for transformation of *S. cerevisiae* was purified using a commercially available kit (Quiagen plasmid purification system, Hybaid) according to the manufacturers instructions.

2.6.3 Isolation of single stranded DNA.

Precipitation solution. 20% Polyethylene glycol (PEG 6000), 2.5 M sodium chloride.

Plasmids derived from pBluescript were transformed into *E. coli* strain XL1-blue and a single colony used to inoculate an overnight culture. 15 µl of

this culture was used to inoculate 5 ml of fresh supplemented LB medium and the culture was grown to early log. phase. 5 µl M13K07 helper phage (Vieira and Messing, 1982) was added to this culture, which was then incubated for 1-2 hours at 37°C. 200 µl of the culture was transferred to 5 ml of fresh LB medium, supplemented with 100 µg/µl ampicillin and 20 µg/µl kanamycin to select for transfected cells and shaken o/n at 37°C.

1.5 ml of this culture was centrifuged to remove the bacterial cells and 1.4 ml of the supernatant transferred to a clean tube and recentrifuged. 1.2 ml supernatant was removed to a fresh tube and 200 µl of precipitation solution added, mixed by inversion and left on ice for 10 minutes. After centrifugation, the pellet was resuspended in 200 µl TE (pH 7.4) and heated to 37°C for 5 minutes. Following extraction with phenol/chloroform at 55°C, single-stranded DNA was precipitated with 0.1 volume 3 M sodium acetate (pH 4.8) and 2.5 volumes ethanol at -70°C for 30 minutes. DNA was pelleted by centrifugation and dissolved in 10 µl TE (final concentration approximately 1 µg/µl).

2.6.4 Estimation of RNA and double-stranded DNA concentration.

RNA concentrations were measured spectrophotometrically using absorbance at O.D.₂₆₀. An O.D.₂₆₀ of 1.0 corresponds to a concentration of 40 µg/ml for RNA. Estimates for DNA concentrations were obtained from ethidium stained samples, electrophoresed together with lambda DNA standards.

2.6.5 Restriction enzyme digestion of DNA

Restriction enzyme digests were performed using the incubation conditions recommended by the manufacturers of the restriction enzymes (Gibco-BRL,

Boehringer-Mannheim). Pretreated RNase A was added to digests of 'miniprep' DNA to remove contaminating RNA.

2.6.6 Agarose gel electrophoresis of DNA

DNA loading buffer. 15 mg Xylene cyanol FF, 15 mg Bromophenol blue dissolved in 5 ml glycerol and 5 ml water.

Agarose gel electrophoresis was routinely used to separate DNA fragments. Agarose concentrations varied from 0.8% to 1.2% (depending on the fragment sizes to be separated) and voltages varied from 10–75 volts. The standard electrophoresis buffer was 1x TAE and standard methods were used for electrophoresis (Sambrook, Fritsch et al. 1989).

Phage λ DNA digested with *Hind*III was used as size marker throughout this study unless mentioned otherwise. Approximately 0.75 μ g were used and the fragment sizes of the size marker are:

23.13 kbp, 9.416 kbp, 6.557 kbp, 4.361 kbp, 2.322 kbp, 2.027 kbp, 0.564 kbp, 0.125 kbp (this last band was usually not visible after staining)

2.6.7 Photography of agarose gels.

Gels were stained in 0.5 μ g/ml ethidium bromide in 1x TAE, and destained in water. Alternatively ethidium bromide was added to the gel and running buffer to a final concentration of 0.5 μ g/ml. Gels were photographed using transmitted short-wave uv light with an orange filter, using an aperture setting of 4.5 and exposure times of 3-7 seconds. The film was Ilford HP-5*plus* (4"x5"). The negative was developed for 6 minutes in 1x Kodak X-ray film developer (LX-24), and fixed in 1x Kodak X-ray fixer

(FX-40) for 4 minutes. The negative was washed in running water for 10 minutes.

2.6.8 Isolation of DNA fragments from agarose gels

Following electrophoresis and staining, the gel was viewed with long wave uv light and the required fragment excised from the gel. DNA was isolated using a commercially available kit (Prep-a-gene system, BioRad) according to the manufacturers instructions. The method yielded up to 85% recovery of DNA.

2.6.9 Cloning Procedures

Cloning of DNA fragments was carried out following standard procedures (Sambrook, Fritsch et al. 1989). The plasmid vector pBluescript (Stratagene) was used routinely for sub-cloning DNA fragments of interest. Successful insertion of a fragment into the cloning site of the polylinker causes insertional inactivation of the β -galactosidase gene. Recombinants are identified following transformation of suitable recipient *E. coli* strains (for example XL-1 blue, DH5 α) because they form white colonies on X-gal medium, compared to blue colonies from cells replicating the uninterrupted vector.

All other vectors used in the current study did not offer a colour selection procedure for the identification of recombinant clones. Cloning into these vectors was routinely carried out with insert and vector fragments cut by two different restriction enzymes (directional cloning). Alternatively the vector fragment was treated with calf intestinal alkaline phosphatase (following procedures recommended by the manufacturer, Boehringer Mannheim) in order to reduce the level of vector self ligation.

Prior to ligation, the DNA fragments were purified by agarose gel electrophoresis followed by band elution. Ligations were carried out for 2-4 hours at RT or o/n at 16°C using DNA ligase and ligation buffer from Gibco-BRL.

2.7 Southern blot analysis

2.7.1 Southern blotting (Southern 1975)

DNA fragments were separated by agarose gel electrophoresis, the gel stained in 0.5 µg/ml ethidium bromide and photographed with a ruler in position alongside the gel to provide the scale for the photograph.

The Southern blot was then carried out as described in (Sambrook, Fritsch et al. 1989) using Hybond-N nylon membrane (Amersham RPN.203N). After the DNA transfer, membranes were air-dried, baked for 1 hour at 80°C and subjected to UV cross linking using a Stratalinker.

2.7.2 Hybridisation procedure

Hybridisations were carried out in a rotary incubator (Hybaid).

Filters were pre-soaked for 1 hour at 65°C in pre-soak solution (5x SSC, 0.5% SDS, 20 mM sodium phosphate buffer (pH7.0), 5x Denhardt's reagent) pre-warmed to hybridisation temperature. Approximately 0.2 ml of solution was allowed for every cm² of nylon filter. The pre-soak solution was removed and the filters were then pre-hybridised at 65°C in pre-hybridisation solution (5x SSC, 0.5% SDS, 20 mM sodium phosphate buffer (pH7.0), 2x Denhardt's, 100 µg/ml denatured salmon sperm) for a minimum of 4 hours.

Radioactive labelling of DNA for Southern Analyses.

DNA hybridisation probes were labelled with $\alpha^{32}\text{P}$ -dCTP (Amersham PB10205, specific activity 3000 μCi , 9.25 MBq) by nick translation using the BRL kit (BRL 530-8160SB) following the manufacturers instructions. Approximately 150 ng of DNA were used per labelling reaction.

To remove unincorporated nucleotides from the radioactively labelled DNA, the sample was passed through a Sephadex G-50 column. The column was prepared in a 145 mm Pasteur pipette plugged with glass wool, and rinsed with 1x TNE before use. The sample was applied to the column and 150 μl aliquots of 1x TNE added. Separate fractions were collected and monitored for incorporation of the radioactive label using a hand held geiger counter. Fractions which contained incorporated label (corresponding to the first peak) were pooled. The labelled probe was denatured by boiling for 10 minutes and then quenched on ice for 5 minutes prior to use in hybridisation.

The radioactively-labelled probe was added to the pre-hybridisation solution and dispersed uniformly. The filters were hybridised overnight with continuous agitation at 65°C.

2.7.4 Washing of filters

After removal of the hybridisation solution, the membranes were washed with continuous agitation and increasing stringency:

Wash 1 (5 minutes at room temperature.) 3x SSC, 5 mM Na phosphate buffer (pH7.0), 0.1% SDS.

Wash 2 (15 minutes at 65°C) 3x SSC, 5 mM Na phosphate buffer (pH7.0), 0.1% SDS.

Wash 3 and 4 (45 minutes each wash at 65°C) 1x SSC, 5 mM Na phosphate buffer (pH7.0), 0.1% SDS.

Wash 5 (for 20-30 mins at 65°C) 0.2x SSC, 5 mM Na phosphate buffer (pH7.0), 0.1% SDS.

2.7.5 Autoradiography.

Filters were exposed to Fuji X-ray film at -70°C with intensifying screens. The film was developed manually in Kodak X-ray developer (LX-24) for 1-6 minutes, depending on the exposure of the film. The developing procedure was stopped in water and fixed in Kodak X-ray Fixer(FX-40) for 4 minutes. The film was washed under running water for 10 minutes.

2.7.6 Re-probing of Hybond-N filters.

After removal of the DNA probe, the Hybond-N filters could be re-probed following the same protocol. The stripping procedure was as detailed by the manufacturer (Amersham International plc). The filters were washed with continuous agitation in 0.4 M sodium hydroxide for 30 minutes at 65°C, followed by further washing for 30 minutes in 0.1x SSC, 0.1% SDS, 0.2 M Tris-HCl (pH7.5). This procedure was repeated if residual radioactivity was detected.

2.8 DNA-sequencing

2.8.1 DNA-sequencing protocol

DNA-sequencing was carried out using the dideoxy chain termination technique (Sanger, Nicklen et al. 1977). Sequencing reactions were

performed using a commercially available kit (T7 DNA polymerase sequencing system, Pharmacia), following the manufacturers instructions both for double and single stranded templates but using only half the recommended amount of $\alpha^{35}\text{SdATP}$ (Amersham).

The sequencing reactions were analysed using the BioRad sequencing gel apparatus (Sequi-Gen Nucleic Acid Sequencing system). Assembly of the apparatus, pouring and running of the gels was carried out according to the manufacturers instructions. Standard gels were 6% acrylamide/bisacrylamide (19:1) (40% premixed stock purchased from Sigma), 6 M urea and 0.5x TBE. The gel solution was filtered through 0.45 μm nitrocellulose prior to use. Polymerisation was initiated by adding 0.1% TEMED and 0.1% 25% ammoniumpersulphate solution and poured immediately. In order to be able to read sequences more than 500 bp away from the cloning site, low percentage gels were prepared. These gels contained 4.5% acrylamide/bisacrylamide (19:1), 6 M urea and 1x TBE.

Once the gel run was complete (3-10 h), the gel was fixed on the glass plate in 10% methanol, 10% acetic acid for 20 minutes and blotted on to pre-wet Whatman 3MM filter paper. The gel was covered with cling film and dried on a BioRad gel drier for 2 hours under vacuum at 80°C. The cling film was removed and the gel exposed to X-ray film at room temperature. The film was developed as detailed in Section 2.7.5.

2.8.2 DNA sequence analysis

All primary DNA sequence data were analysed using the GCG sequence analysis package version 7.3 (Group 1991). Similarity searches were carried out using the FASTA search facility of the GCG package and the BLAST algorithm as provided via a telnet service at the NCBI

supercomputer (Altschul, Gish et al. 1990). Protein secondary structure predictions were obtained using the PHD Predict Protein mailserver at PredictProtein@EMBL-Heidelberg.DE (Rost, Schneider et al. 1993).

2.9 RNA isolation and manipulation.

2.9.1 RNA Extraction from *Coprinus cinereus* using the Guanidinium Thiocyanate method. (Glisin, Crkvenjakov et al. 1974) (Sambrook, Fritsch et al. 1989), with modifications suggested by (Chirgwin, Przybyla et al. 1979)

(AnalaR water was used for the preparation of all solutions)

<u>Homogenisation Buffer</u>	<u>Guanidinium Hydrochloride Solution</u>
5.5 M Guanidinium thiocyanate	
0.025 M Sodium-citrate (pH7.0)	7.5 M Guanidinium hydrochloride
0.1% Antifoam A (Sigma A5758)	0.05 M Sodium-citrate (pH7.0)
0.175% N-lauryl sarcosine	
0.2 M β -mercaptoethanol.	

Ultracentrifugation solution. 5.7 M Caesium chloride, 0.1 M sodium-EDTA

Preparation of utensils

All glassware and centrifuge tubes (polyallomer ultracentrifuge tubes and 50 ml oakridge centrifuge tubes) for the preparation of RNA were acid washed (10-15% Hydrochloric acid) and autoclaved. Pasteur pipettes were baked at 180°C

RNA Protocol

The mycelium from 200 ml liquid cultures was harvested by filtration, washed twice with distilled water, and excess liquid removed by blotting with filter paper. The mycelium was homogenised with a polytron

homogeniser (Kinematica PT10/35), using 10 ml homogenisation buffer per 10 g fresh weight mycelium. The slurry was centrifuged at 13K for 20 minutes at room temperature and the supernatant layered onto 5.7 M caesium chloride, 0.1 M sodium-EDTA in the following proportions:-

Ultracentrifuge rotor	V o l u m e supernatant	o f Volume of caesium chloride
TL-100	2.0 ml	1.0 ml

After ultracentrifugation at 65K (TL-100) for 6 hours, the supernatant was carefully removed without dislodging the pellet and the tube inverted to drain. The sides of the tube were dried with tissue paper and the pellet resuspended in 0.85 ml guanidinium hydrochloride solution. After thorough aspiration to resuspend the pellet, the solution was transferred to a 50 ml oakridge tube containing 1 ml guanidinium hydrochloride. 0.1 ml 1 M acetic acid and 1 ml 95% ethanol was added, the contents mixed by vortexing and the RNA was precipitated for 2 hours at -20°C. The RNA was pelleted at 14 K for 30 minutes at 4°C and the supernatant discarded. The tube was inverted to drain, the pellet partially air-dried and resuspended in 1.8 ml water. 0.2 ml 1 M sodium acetate pH 5.0 and 5 ml 95% cold ethanol were added to the RNA solution, mixed and the RNA precipitated for at least 2 hours at -20°C. After centrifugation at 15K, the supernatant was discarded and the pellet rinsed with cold 70% ethanol and extensively air dried, then dissolved in 0.5-1.0 ml water. RNA was stored at -70°C.

2.9.2 Isolation of poly (A)⁺ RNA.

<u>5X loading buffer</u>	<u>Elution buffer</u>
2.5 M potassium chloride	0.01 M TrisHCl (pH7.5)
0.05 M TrisHCl (pH 7.5)	

Preparation of Oligo d(T) cellulose affinity columns.

0.25 g of oligo d(T) cellulose (Pharmacia) was equilibrated overnight in 1.5-2.0 ml 1x loading buffer at 4°C. The columns were prepared with a bed volume of 0.5 ml, in 1 ml pipette tips, plugged with autoclaved siliconised glass wool. Spring clips were used to regulate the flow rate through the column.

The columns were sterilised prior to use by washing with 5-10 ml 0.1 M sodium hydroxide, followed by equilibration for 20 minutes with 0.1 M sodium hydroxide at room temperature. The column was washed with 5 ml sterile water and then with 6 ml 1X loading buffer until the pH of the eluate was less than pH 8.0. The flow rate was adjusted to 1 drop every 20 seconds.

Isolation of poly(A)⁺ RNA

RNA (2-5 mg) was denatured at 65°C for 15 minutes, quenched on ice and applied to the column in a volume of 1 ml with 0.25 ml 5x loading buffer. The RNA solution was kept on ice between loadings and reapplied to the column five times. The column was washed with 0.5 ml 1X loading buffer which was combined with the unbound RNA solution (poly(A)⁻ RNA). The column was washed with 5-10 ml ice cold 1x loading buffer to remove the unbound poly(A)⁻ RNA. The poly(A)⁺ RNA was eluted from the column with elution buffer, prewarmed to 45°C and six 0.5 ml fractions collected. The poly (A)⁺ RNA fractions were identified by UV fluorescence; 3 ml of each fraction was combined with 20 ml ethidium bromide (1 mg/ml), spotted onto clingfilm and visualised on a transilluminator using a water control (water and ethidium bromide) for comparison. Fractions containing Poly(A)⁺ RNA were pooled and applied to a second equilibrated

column after denaturation. This removed contaminating poly(A)⁻ RNA. The fractions containing poly(A)⁺ RNA were pooled.

Poly (A)⁺ RNA and poly (A)⁻ RNA were precipitated overnight at -20° C with the addition of 0.1 volume ice cold 3 M potassium acetate (pH 5.4) and 2.5 volumes ice cold ethanol. RNA was pelleted by centrifugation at 10,000x g for 30 minutes, the supernatant discarded and the pellet washed with ice cold 70% ethanol. RNA was resuspended in water: for 2 mg total RNA input, 35 ml for poly(A)⁺ RNA and 500 ml for poly (A)⁻ RNA were used.

The concentration of the RNA was assessed spectrophotometrically and samples run on a denaturing gel to check for degradation (See Section 2.10.3).

The columns were sterilised for re-use by washing with 5 ml 0.1 M sodium hydroxide and 5-10 ml 1x loading buffer containing 0.02% sodium azide. After sealing with parafilm, the columns were stored submersed in loading buffer + 0.02% sodium azide at 4°C.

2.9.3 RNA agarose gel electrophoresis.

5x MOPS buffer 0.2 M MOPS (3-morpholinopropanesulfonic acid), 50 mM sodium acetate, 5 mM EDTA. The pH was adjusted to 7.0 with 1 M sodium hydroxide.

10x RNA loading buffer 50% (v/v) glycerol, 1 mM EDTA, 0.4% (w/v) bromophenol blue, 0.4% (w/v) xylene cyanol. Stored at 4°C.

For a 1% gel, agarose was autoclaved in half the final volume of water, cooled to 60°C and 0.2 final volume 5X MOPS, and 0.175 final volume

formaldehyde added. The volume was adjusted with water and the gel was poured in an ethanol-rinsed gel tank.

RNA was loaded in the following solution:-

22.5% RNA in water (5-10 μ g)

10.0% 5x MOPS

17.5% formaldehyde

50.0% deionized formamide

The solution was incubated at 55°C for 15 minutes, quenched on ice for 5 minutes and 0.1 volume of 10x RNA loading buffer added prior to loading. The gel was run in 1x MOPS, pre-running for 15 minutes before loading. An RNA size marker (0.24-9.5 kb) (Gibco-BRL S620SA), pre-treated identically to the RNA samples was also loaded onto the gel. The size marker was visualised by uv fluorescence after staining in 1 mg/ml ethidium bromide solution.

2.10 Northern blot analysis.

2.10.1 Northern blotting.

To remove the formaldehyde from the gel prior to blotting, the denaturing RNA gel was soaked in water for an hour, changing the water every 10 minutes. The molecular weight standard was cut off, stained with ethidium bromide and photographed with a ruler for future reference.

The gel was soaked twice in 20x SSC for 20 minutes. A Northern blot was set up as for a Southern blot (Sambrook, Fritsch et al. 1989). After RNA transfer, the membrane was briefly rinsed in 3x SSC and air-dried for 1 hour, baked for 1 hour at 80°C and subjected to crosslinking using a Stratalinker.

2.10.2 Northern hybridisation.

Hybridisation Solution

5x SSPE, 50% (v/v) deionized formamide, 5x Denhardt's Reagent,
0.5% SDS, 400 mg/ml denatured salmon sperm DNA

The filter was prehybridised in 7.5-10 ml hybridisation solution for a minimum of 4 hours in small bags at 42°C, shaking at <100 rpm. The probe was added directly to the bag and hybridised overnight at 42°C. Radioactive probes were prepared using a commercially available random prime labelling system (Boehringer Mannheim) and $\alpha^{32}\text{P}$ dCTP (Amersham) following the manufacturers instructions. Unincorporated nucleotides were removed as described above for nick-translated probes by Sephadex G-50 size exclusion chromatography. The labelled probe was boiled and quenched on ice before adding to the hybridisation bag.

2.10.3 Washing of filters.

The filters were washed in the following solutions:-

Wash 1 and 2 (For 30 minutes at 42°C) 4x SSC, 25 mM Sodium phosphate (pH7.0), 0.1% SDS.

Wash 3 (For 30 minutes at 42°C) 1x SSPE, 0.1% SDS.

Wash 4 and 5 (For 5 minutes at 50°C) 0.1x SSPE, 0.1% SDS.

Wash 6 (5 minutes at 50°C) 0.1x SSPE, 0.1% SDS

The standard methods of autoradiography were used (See Section 2.8.6).

2.10.4 Stripping and re-probing of Northern filters.

The Hybond-N filters were stripped of the hybridised labelled probe as detailed by the manufacturer (Amersham International plc). The filters were washed three times for 45 minutes at 65°C with continuous agitation in 0.005 M Tris-HCl (pH8.0), 0.002 M Sodium EDTA (pH8.0) and 0.1 x Denhardt's reagent. Filters could then be reprobed.

2.11 Construction of a cDNA library

A cDNA library was constructed using the Stratagene ZAP-cDNA™ synthesis kit and GigapackII Gold Packaging extracts (Stratagene).

2.11.1 Synthesis of cDNA

The cDNA was synthesised according to the kit manufacturers instructions with some modifications. 5 µg of poly(A)+ RNA were used as input into the first strand reaction mixture. A 5µl aliquot of the first strand reaction mix was removed and 0.5 µl of $\alpha^{32}\text{PdATP}$ (3000 Ci/mmole instead of 800 Ci/mmole as suggested in the kit manual) was added to this. The 2nd strand synthesis was carried out with the addition of 2 µl $\alpha^{32}\text{PdATP}$ (again at the higher specific activity).

After the 2nd strand synthesis, an 8 µl aliquot was removed. 2 µl of this together with 1 µl of 1st strand control reaction (5 µl aliquot that had added radiolabel) were spotted onto glass fibre filter disks and air dried for 15 minutes. The filter disks including a blank were transferred into scintillation vials and Cerenkov counted (Cerenkov 1934) in a scintillation counter (^3H channel for 4 min) (Simon 1974).

The filter disks were then washed with cold 5% trichloro acetic acid (5 washes of 3 minutes), dried at 80°C for 20 minutes, allowed to cool to room temperature and Cerenkov counted as above. The data obtained made it possible to calculate the yield of 1st and 2nd strand cDNA reactions. This is illustrated for the 1st strand reaction below:

1st strand:

0.3 µl 10 mM dATP (unlabelled) in the 5µl set aside control reaction equals 3×10^{-9} moles of dATP.

ratio of incorporation of dATP was obtained by dividing the two cpm readings obtained from the Cerenkov counting (incorporated = after washing, total = before washing):

$$(\text{incorporated})/(\text{total}) = 14854/788700 = 0.018833$$

It follows that moles of dATP incorporated is:

$$0.018833 \times 3 \times 10^{-9} \text{ moles} = 5.65 \times 10^{-11} \text{ moles}$$

This number multiplied by four gives the amount (in moles) of all four nucleotides incorporated during the reaction:

$$\text{dNTP}_{\text{inc}} = 4 \times 5.65 \times 10^{-11} \text{ moles} = 2.26 \times 10^{-10} \text{ moles}$$

The control reaction was only 10% of the volume of the 1st strand reaction, therefore the moles of dNTP incorporated in the 1st strand reaction is:

$$2.26 \times 10^{-9} \text{ moles}$$

Taking 360 g as the average weight of 1 mole of dNTP, it follows that $2.26 \times 10^{-9} \text{ moles} \times 360 \text{ g/mole} = 8 \times 10^{-7} \text{ g} = 0.8 \text{ µg}$ of 1st strand cDNA were synthesised.

The amount of 2nd strand cDNA produced was estimated following an analogous calculation.

The quality of the 1st and 2nd strand cDNA was assessed by alkaline gel electrophoresis. The alkaline agarose gel and loading buffer were prepared as described in the appendix of the lambda ZAP manual. As size marker, phage lambda DNA digested with *Hind*III was end labelled with ³²PdATP. The digested lambda DNA was incubated for 2 hours at RT in the following reaction solution:

- 12 µl lambda *Hind*III
- 2 µl buffer H (Boehringer Mannheim restriction enzyme buffer)
- 3 µl solution A (Gibco-BRL nick translation kit, containing dCTP, dGTP and dTTP at 0.2 mM each)
- 1 µl (2u) Klenow enzyme (Boehringer Mannheim)
- 2 µl ³²PdATP (Amersham)

6 µl of this size marker was loaded onto the gel alongside 4 µl of the set aside (5 µl aliquot) 1st strand control reaction (5 µl aliquot) and the residual 6 µl of the 2nd strand reaction sample. The gel was run for 3.5 hours at 100 mAmp, subsequently fixed in three changes of 7% TCA, soaked o/n in water and dried between paper towels. The gel was then wrapped in cling film and exposed to X-ray film, in order to show the size distribution of the synthesised cDNAs.

2.11.2 Cloning of cDNA

The cDNA was cloned into the λ ZAP arms according to the method described in the manual with one modification:

After the spin column for size separation, 6 µl aliquots of each fraction were removed to analyse the amount and quality of cDNA in each fraction. 1 µl was used for Cerenkov counting as described above and this allowed a calculation of the amount of cDNA in each fraction according to the following formula:

$$\frac{\text{cpm/frac.} \times \text{vol/frac.}}{0.5 \times \text{cpm}(\text{total}) \times \text{vol}(\text{total})} \times 2000 \text{ ng} = \text{ng cDNA per fraction}$$

cpm(total) was the counts of incorporated nucleotides after the 2nd strand cDNA synthesis. This was multiplied by 0.5 to adjust for the fact a 2 µl aliquot of the total reaction volume of 400 µl had been used for Cerenkov counting after the 2nd strand synthesis.

The remaining 5 µl were analysed by agarose gel electrophoresis. After a standard gel run, the size marker track was removed, stained and photographed alongside a ruler for future reference. The remaining part of the gel was dried between paper towels, wrapped in cling film and exposed to X-ray film in order to be able to monitor the size distribution of the cDNAs in the individual fractions.

2.11.3 Maintenance of the cDNA library

The ligated cDNA was packaged using the GigapackII Gold packaging extracts (Stratagene) as recommended by the manufacturer. The primary library was amplified once (as recommended) and stored at 4°C over chloroform. This phage stock could be used directly for mass "dropping out experiments", i.e. mass release of pBluescript cDNA clones from the lambda ZAP phages. These experiments were carried out following the technique described in the manual using the ExAssist helper phage and the *E. coli* strain SOLR.

The average insert size of the cDNA library was determined using two different approaches. Firstly, ten randomly chosen plaques were cored out and the cDNAs released as pBluescript clones. Secondly, a mass release from 4×10^5 pfu of the amplified phage stock 1-1 was carried out. Plasmid

minipreps were prepared and allowed an estimation of the average insert size to be 1.1 kbp. Surprisingly, two of these randomly chosen cDNA clones were not cut by *EcoRI*. It was later discovered through sequencing of positive cDNA clones, that a high proportion of the cDNA clones in this library had lost one base in the *EcoRI* cloning site.

2.11.4 Screening of the cDNA library

For screening purposes, the amplified library was plated onto 140 mm petri dishes with approximately 9,000 pfu per plate using the bacterial strain XL1Blue. The plates were incubated at 37°C for up to 8 hours until the plaques were roughly 1 mm in diameter. After chilling the plates at 4°C for at least 1 hour, two replica filter lifts were taken using Hybond N filter circles (Amersham), following the instructions of the membrane manufacturers. The membranes were air dried and baked at 80°C for 1 hour. The DNA was covalently linked to the nylon membrane using a Stratalinker (Stratagene).

The library filters were hybridised in polythene bags at 65°C as for Southern blot hybridisations (section 2.7.2) using isolated DNA fragments as hybridisation probes. Positive plaques were identified by corresponding spots on replica filters. These were isolated and placed in 0.5 ml of SM buffer with a drop of chloroform, left o/n at 4°C and then used for the second round screen.

For the second round screen, the phage suspension was suitably diluted, plated onto 90 mm NZY plates and incubated overnight. Single filter lifts were taken from plates with approximately 100 plaques. Hybridisation to the initial probe identified individual positive plaques. These were placed in

0.5 ml SM buffer with a drop of chloroform and stored at 4°C. pBluescript cDNA clones were isolated for further analysis.

2.12 Screening of the LambdaGEM 11 genomic library

2.12.1 Plating and screening the library.

The library was plated onto four 150 mm LB agar plates at a density of 3,000-4,000 pfu/plate using KW251 phage competent cells prepared as described in the λ GEM11 manual (Promega). Plates were incubated overnight at 37°C until the plaques were well-defined and approximately 2 mm in diameter. After one hour at 4°C, two replica filter lifts were taken from each plate and following standard fixation procedures screened for phages containing DNA sequences of the A6 factor (method as for cDNA library, section 2.11).

2.12.2 Small scale isolation of λ GEM 11 phage DNA

The titre of the phage stock was assessed by serial dilution. Approximately 5×10^3 pfu in a volume of 100 μ l was adsorbed on to fresh KW251 competent plating cells for 15 minutes and plated on 90 mm LB plates. After incubation o/n at 37°C, the plates were chilled at 4°C for 1 hour. 5 ml SM buffer was added to the plates which were then gently shaken at 4°C on an orbital shaker for 2-4 hours. The SM-lysate was removed into an autoclaved centrifuge tube and the plate rinsed with a further 2 ml SM buffer, gently shaken for a further 15 minutes and the lysate collected. After centrifugation for 10 minutes at 4000 rpm at 4°C, the supernatant was removed and stored at 4 °C. This high titre phage stock was used for phage DNA purification.

2.12.3 Phage DNA purification.

DEAE cellulose resin. 50 g DEAE cellulose (Whatman, DE52) was placed in 0.05 M hydrochloric acid and stirred (pH <4.5). The pH was adjusted to 7.0 with 1 M sodium hydroxide, and the resin left to settle. The supernatant was decanted off, the resin resuspended in 250 ml LB medium and allowed to settle. The supernatant was again decanted and a further 250 ml LB added. After decanting a third time, the resin was resuspended in 0.25 volumes LB medium with the addition of 0.1% (w/v) sodium azide and stored at 4°C.

4 ml sterile 20%(w/v) PEG/20% (w/v) NaCl solution was added to 4 ml phage plate lysate stock, mixed and incubated for 1 hour in ice/water. Following centrifugation at 3,000 rpm for 20 minutes in a bench top centrifuge, the supernatant was decanted. The inside of the tube was wiped with a tissue to remove all traces of PEG and the pellet resuspended in 750 µl LB, and transferred to an Eppendorf tube. An equal volume (750 µl) of DE52 in LB was added to the tube, mixed by inversion 20-30 times and centrifuged for 5 minutes. The supernatant was transferred to a clean tube and re-centrifuged. To each 1 ml of supernatant, 17.5 µl of a 0.1 mg/ml solution of proteinase K and 42.5 µl of 10% SDS was added, mixed and left at room temperature for 5 minutes. 173 µl of 3 M potassium acetate was added and a precipitate formed, which dissolved after incubation at 88°C for 20 minutes. After cooling on ice for 10 minutes, the tube was centrifuged for 10 minutes and the supernatant transferred to two fresh microcentrifuge tubes. An equal volume of cold isopropanol was added and the sample left at -70 °C for 10 minutes and then warmed to room temperature. After centrifugation for 10 minutes to pellet the phage DNA, the pellet was drained and air dried. The combined phage pellet was resuspended in 40 µl TE (pH 8.0).

2.13 Gel electrophoresis of proteins

2.13.1 SDS-PAGE minigels

Electrophoretic separation of proteins was achieved by SDS-PAGE using the discontinuous buffer system of (Laemmli 1970). Minigels were prepared and run using the AE.6450 gel tank (ATTO corporation) following the manufacturers instructions. Separating gels of 8% to 14% were prepared with an acrylamide stock solution of 36 acrylamide : 1 bisacrylamide, 20% gels had a ratio of 280 acrylamide : 1 bisacrylamide. Stacking gels were prepared at 3.75% acrylamide (36 : 1).

Molecular weight standards were obtained from SIGMA and Amersham and treated according to the manufacturers instructions.

SIGMA	SIGMA	Amersham
MW-SDS-70L (kd)	MW-SDS-200 (kd)	Rainbow™ (kd)
66	205	200
45	116	97.4
36	97	69
29	66	46
24	45	30
20	29	21.5
14		14.3

Staining and drying minigels

The proteins were fixed in the gel matrix by immersing the gel. in 20% TCA for 20 min and afterwards rinsing in dH₂O for 2 min.. Gels were then stained using Coomassie brilliant blue (R250) following standard procedures (Sambrook, Fritsch et al. 1989). For drying, gels were placed between 2 prewetted sheets of cellophane and dried under vacuum on a BioRad gel dryer at 80°C for 1 to 2 hours. Alternatively gels were placed on a piece of 3MM paper (Whatman), covered with Saran Wrap and dried as above.

Fluorography of SDS-PAGE minigels

Gels with radioactively labelled proteins were analysed by fluorography. Immediately after the gel run, the proteins were fixed in the gel matrix by immersing the gel in 10% acetic acid/10% methanol for at least 30 minutes. The gel was then soaked in AmplifyTM (Amersham) for 20 to 30 minutes, dried on 3 MM paper as described above and exposed to X-ray film at -70°C with an intensifying screen. The prestained RainbowTM size markers allowed estimation of the sizes of any bands appearing on the autorad film.

2.13.2 Preparative SDS-PAGE

Preparative SDS-PAGE was carried out in order to purify recombinantly produced protein for immunisation. The protein samples were separated on a 6% (36 acrylamide : 1 bisacrylamide) gel using the Hoefer SE 600 gel tank (16 cm x 18 cm gel).

The gel was assembled and electrophoresed according to the manufacturers instructions. The proteins to be purified were in the form of

insoluble inclusion bodies. These were solubilised by adding an equal volume of 2x SDS-Gel sample buffer and boiling for 10 minutes. 400 µl solubilised protein was loaded into 4 wells of the gel. Adjacent wells contained a 10 µl aliquot of inclusion body preparation and size markers. After electrophoresis, these two control tracks were cut off and stained with Coomassie brilliant blue. The remainder of the gel was immersed in ice cold 3 M potassium chloride for three minutes which led to the formation of a white precipitate at regions of very high protein concentration. The two parts of the gel were realigned to allow precise excision of a gel strip containing the recombinant protein. The gel slice was dialysed against two changes of PBS at 4°C and then macerated by cutting with a razor blade and then squeezing through 19 and 21 gauge needles. The material was stored at -70°C. The purity of the gel-purified protein was assessed by taking a small aliquot and boiling this in 1x SDS sample buffer prior to SDS-PAGE analysis.

2.14 Expression of genes in pGEX vectors (Smith and Johnson 1988)

2.14.1 Small scale expression

5 ml o/n cultures of E.coli XL1Blue containing the pGEX vector/expression construct were grown overnight at 37°C in LB medium containing 100 µg/ml ampicillin. 0.5 ml of this culture was inoculated into 5 ml of fresh medium and grown at 37°C for 1 hour. After removing an aliquot (uninduced sample), IPTG was added to a final concentration of 1 mM and the cultures were further incubated at 37°C. Aliquots could now be removed at different times to measure the time course of the recombinant protein production (induced samples).

For rapid analysis, 0.5 ml of cell culture were pelleted in a microcentrifuge for 1 min at full speed. The cells were resuspended in 80 µl 1x loading buffer and boiled in a water bath for 3 min. 5 µl of this crude lysate was usually enough to monitor the production of recombinant protein on an SDS-PAGE minigel.

2.14.2 Large scale isolation of HDX3 GST fusion protein

800 ml LB-Amp were inoculated with an 80 ml o/n culture of *E. coli* XL1Blue containing the pHDX3 plasmid. After 1 h at 37°C, IPTG was added to a final concentration of 1 mM and the culture was grown for another 4 h at 37°C with vigorous shaking. Cells were harvested by centrifugation, washed in ice-cold MTPBS, lysed by sonication (2 x 30 sec on ice 6 mm probe, setting 5 with 40% duty cycle) and the insoluble fraction was collected by centrifugation. The insoluble protein was solubilised by adding urea to a final concentration of 6 M. After 10 min at 4°C, cell debris was removed by centrifugation and the supernatant was dialysed against two changes of PBS, containing 1% triton X-100, 1% Tween 20 and 10 mM DTT at 4°C as suggested in (Smith and Johnson 1988). HDX3 protein solubilised in this way could not be purified using the affinity matrix glutathione agarose (Sigma).

2.14.3 Large scale inclusion body preparation

20 ml of an o/n culture of *E. coli* XL1Blue carrying a pGEX expression vector construct were inoculated into a fresh 200 ml LB culture containing ampicillin. After growth for 1 hour at 37°C, IPTG was added to a final concentration of 1 mM and the cells were grown for 4 more hours with vigorous shaking. The cells were then harvested by centrifugation (2.5 K, 15 min, 4°C) and resuspended in 10 ml of lysis buffer (50 mM Tris pH8, 25%

sucrose, 1 mM EDTA, 2 mg/ml lysozyme). After adding 15 ml of detergent buffer (0.2 M NaCl, 20 mM Tris pH7.5, 2 mM EDTA, 1% deoxycholate, 1% NP40), the cells were incubated for 15 min on ice. $MgCl_2$ was added to 10 mM and $MnCl_2$ to 1 mM and the degradation of DNA was initiated by adding 50 μ l of 10 mg/ml DNase solution. After a further incubation on ice for 30 min, the inclusion bodies were collected by centrifugation (5 K, 10 min, 4°C), washed 4 times with triton wash solution (0.5% triton X-100, 1 mM EDTA) and finally resuspended in 1 ml of PBS.

SDS-PAGE allowed estimation of the obtained amount of inclusion body protein by using the molecular weight marker bands as standards. An approximate yield of 10 mg of recombinant protein was obtained per 200 ml culture.

2.14.4 Large scale isolation of soluble GST protein

A 200 ml culture of XLIBLue carrying the pGEX2T plasmid was grown, induced and harvested as described for the inclusion body preparation. The cell pellet was resuspended in 10 ml MTPBS and the cells were disrupted by sonication (2 x 30 sec on ice, 6 mm probe, setting 5 with 40% duty cycle). After adding triton X-100 to 1%, the cell debris was pelleted (10,000 g, 5 min, 4°C) and aliquots of the supernatant frozen at -70°C.

2.14.5 Expression in the presence of sorbitol and betaine

In order to achieve at least partial solubility of the pGEX fusion proteins, the growth conditions were altered. The cells were grown at 28°C and sorbitol and betaine were added growth medium to a final concentration of

0.6 M and 2.5 mM respectively as suggested by (Blackwell and Horgan 1991).

A 5 ml culture of cells containing the pGEX expression vector construct was grown for 24 hours at 28°C. This was used to inoculate a 250 ml culture which was grown o/n at 28°C. After addition of IPTG to 1 mM, the cells were grown for a further 2 hours at 28°C. The cells were harvested, sonicated and stored as described.

2.14.6 Thrombin cleavage of fusion proteins

The soluble fusion proteins were cleaved with thrombin. This proteolytic cleavage of the fusion proteins was carried out while the proteins were adsorbed to the affinity matrix glutathione-S-sepharose. 100 µl of glutathione-S-sepharose beads were washed three times with 900 µl MTPBS. After three further washes with 900 µl MTPBS/1% triton X-100, the beads were incubated for 15 min at RT with 300 µl crude *E. coli* protein extract (2.14.4) containing the GST fusion protein. The beads were pelleted, the supernatant removed and the resin washed three times with 900 µl MTPBS. The resin was washed with 900 µl 50 mM Tris pH8/150 mM NaCl/2.5 mM CaCl₂ and resuspended in 100 µl of the same solution. After the addition of 100 ng thrombin (SIGMA T6634, 10 ng/µl stock frozen at -20°C in aliquots), the cleavage reaction was left at RT for 15 to 45 minutes. The resin was pelleted and the supernatant kept for SDS-PAGE analysis. The resin was washed with 900 µl MTPBS and analysed by SDS-PAGE.

2.15 Expression of genes in the pT7-7 vector

2.15.1 Expression using a two plasmid system

In this system, the T7 RNA polymerase is provided by the plasmid pGP1-7, whereas the gene to be expressed is cloned into the multiple cloning site of the plasmid pT7-7 (Tabor and Richardson 1985). The experiments were carried out using the *E. coli* strain WM1704, since this strain has a deletion for the Ion protease complex and this should facilitate the expression of recombinant proteins.

WM1704 cells carrying the two plasmids were grown at 30°C in enriched medium (2% tryptone, 1% yeast extract, 0.5% NaCl, 0.2% glycerol, 50 mM potassium phosphate pH7.2, 50 µg/ml ampicillin, 50 µg/ml kanamycin) to an OD₆₀₀ of 1.5. The cultures were now grown at 42°C for 25 minutes. Rifampicin was added to a final concentration of 100 µg/ml and the cells were grown for a further 2 hours at 37°C. The cells were then analysed by SDS-PAGE.

2.15.2 Expression using *E. coli* BL21(DE3) strains

The *E. coli* strain BL21(DE3) and its derivatives BL21(DE3)LysE and BL21(DE3)LysS carry the T7 RNA polymerase gene inducible through IPTG on a prophage (Studier, Rosenberg et al. 1990). All three strains were used in expression experiments with pT7-7 expression vector constructs.

5 ml o/n cultures of the respective strain containing the expression construct were grown at 37°C in LB medium containing 100 µg/ml ampicillin. 0.5 ml of this culture were inoculated into 5 ml of fresh

medium and grown at 37°C for 1 hour. After removing an aliquot (uninduced sample), IPTG was added to a final concentration of 1 mM and the cultures were further incubated at 37°C. Aliquots could now be removed to measure the time course of the production of any recombinant protein (induced samples). Samples were analysed by SDS-PAGE.

2.17 Preparation of antisera

The macerated gel slices described in section 2.13 were used directly for immunisation of rabbits after adding an equal volume of Freund's adjuvant. All immunisations and bleeds were kindly carried out by Mrs J. Newell from the MRC Immunochemistry Unit, Dep. of Biochemistry, Oxford University. Sera were prepared following standard procedures (Harlow and Lane 1988). The blood was allowed to clot at RT for 2-4 hours and then left o/n at 4°C. The serum was carefully removed after centrifugation and stored in aliquots at either -20°C or -70°C.

2.17 Western blot analysis

After SDS-PAGE, proteins were electrophoretically transferred to Immobilon PVDF membrane using the Hoefer TE42 Transphor Unit and a glycine/Tris/methanol buffer system (Towbin, Staehlin et al. 1979).

For detection of proteins using colour substrates, Western blots were blocked o/n at 4°C in TBS/5% marvel. After three further 15 minute washes in TBST (TBS/0.1% Tween 20), the blots were incubated o/n at 4°C in TBS/5% marvel + antiserum. After three washes in TBST, the blots were incubated for 2 h at RT in TBS/5% marvel + antirabbit IgG antibody phosphatase conjugate (Sigma. Prod. No. A-8025 at 1 in 500 dilution). After three more

15 min washes with TBST, the western blots were developed in substrate solution (10 mg NaphtholAs-MX Phosphate dissolved in 1 ml DMF + 49 ml 0.1 M Tris pH 8.2 and 50 mg Fast Blue BB salt).

For chemiluminescent detection of proteins, the membranes were blocked o/n in PBS/5% marvel/0.1% Triton X-100 at 4°C. All incubations and washes were as indicated in the kit manual (Amersham ECL Western blot detection system). The membranes were exposed to X-Ray film for 5-100 sec, and the film was developed as described before for Southern blot analysis.

2.18 *In vitro* transcription/translation

Prior to *in vitro* transcription, plasmid DNA was linearised at the 3' end of the gene (or part of a gene) to be translated. The cleaved DNA was purified using a commercially available kit following the manufacturers instructions (GeneClean II DNA purification system, BIO 101).

The *In vitro* transcription/translation was then carried out following the Promega protocol (1991). All reagents were purchased from Promega, with the exception of m⁷G(5')ppp(5')G, which was from Boehringer Mannheim and ³⁵S methionine, which was obtained from NEN/DuPont (NEG-009A). *In vitro* translated proteins could be stored at -70°C.

2.19 Crosslinking of *in vitro* translated proteins

<u>5 x SDS-gel sample buffer:</u>	4 ml dH ₂ O, 1 ml 0.5 M Tris pH6.8
	0.8 ml glycerol, 1.6 ml 10% SDS
	0.4 ml β mercaptoethanol
	0.2 ml 0.05% bromophenol blue

Chemical crosslinking of *in vitro* translated proteins was carried out using the two crosslinking agents DMSI and glutaraldehyde. For both crosslinking agents, control reactions had to be carried out without the reactive chemical in the crosslinking assay.

a) Crosslinking with DMSI (dimethyl suberimidate)

3 µl of the *in vitro* translation mixture were diluted into 100 µl 0.2 M potassium phosphate buffer pH8.5/12 mM DMSI (without DMSI for the negative control) and incubated for 30 minutes at RT. Proteins were then precipitated through the addition of an equal volume of 20% TCA, pelleted for 30 seconds and washed with cold 80% acetone. Proteins were redissolved in 16 µl 5% SDS, and after addition of 4 µl of 5 x SDS-gel sample buffer, analysed by SDS-PAGE.

b) Crosslinking with glutaraldehyde

3 µl of the *in vitro* translation reaction were diluted into 200 µl of 0.1 M NaCl/0.001% glutaraldehyde (without glutaraldehyde for the negative controls) and left at RT for 30 min. After precipitation with an equal volume of 20% TCA, the samples were treated as for the DMSI crosslinking.

2.20 Immunoprecipitation of *in vitro* translated protein

Immunoprecipitation was carried out using the immunoprecipitin reagent from Gibco-BRL (formalin treated *S. aureus* cells, cat. no. 5609321SA), pretreated according to the manufacturers instructions. 5 µl of a b1-1 *in vitro* translation mix were immunoprecipitated following the protocol in (Kessler 1981) with minor modifications. For co-immunoprecipitation of

two different proteins, the respective RNAs were mixed prior to the translation reaction, so that they were co-translated.

The translation mix was first incubated with 5 µl pre-serum and immunoprecipitin and any possible non specific reaction products were removed by centrifugation. The supernatant was then incubated with 5 µl of antisera and left for 2 h at RT. Antigen-antibody complexes were recovered by centrifugation following a further incubation for 30 min on ice with 100 µl of immunoprecipitin. The supernatant was concentrated by trichloro acetic acid precipitation. Aliquots of the preimmune pellet, the immune pellet and the supernatant were analysed by SDS-PAGE.

2.21 Preparation of *C. cinereus* crude protein extracts

Five day small scale liquid cultures of *C. cinereus* were harvested by filtration and ground to a fine powder in liquid nitrogen. This powder could be stored at -70°C. For SDS-PAGE analysis, approximately 100 mg were resuspended in 300 µl SDS-gel sample buffer, boiled for 5 minutes, cooled on ice, sonicated for 30 seconds (40% duty cycle, microprobe setting) and then boiled for a further 5 minutes. Before loading, the samples were centrifuged for 10 minutes and 5-20 µl of the supernatant was used to load each track on an SDS-gel.

2.23 Preclearing of antisera

Approximately 200 mg of protein powder of *C. cinereus* strain NA2 were extracted on ice with intermittent vortexing for 15 min in 1 ml PBS/1 mM EDTA/0.5 mM PMSF/6 mM βMercaptoethanol. The cell debris was removed by centrifugation and the supernatant together with 200 µl of

crude *E. coli* protein extract containing the GST protein was added to 8 ml of PBS. A nitrocellulose filter circle was incubated for 1 h at RT in this solution. The filter was gently washed with PBS and subsequently blocked o/n at 4°C in PBS/5% marvel. The filter was then washed three times with PBS and incubated with the respective antisera diluted into PBS/5% marvel. After 1 hour at room temperature, the precleared antisera could be used directly for Western blot analysis and immunofluorescence.

2.23 Immunofluorescence with fungal mycelia

Mowiol Mountant: Mix 3 g glycerol and 2.4 g of Mowiol (Mowiol 4-88, Hoechst, Frankfurt) and stir at RT for 1 hour. Then add 6 ml of dH₂O and stir for 2 more hours. After adding 12 ml of 0.2 M Tris-HCl (pH8.5) and incubating at 50°C for 10 min, spin at 5,000 g for 15 min. Add 2.5% DABCO (1,4-diazobicyclo[2,2,2]octane) and 1µl/ml DAPI (4',6'-Diamidino-2-Phenylindole) and store in aliquots at -20°C.

Immunofluorescence with *C. cinereus* mycelium was carried out according to the following procedure kindly provided by Dr. M. Dewey, Dep. of Plant Sciences, University of Oxford. Teflon coated glass slides were autoclaved, placed in the bottom of a petri dish and coated with solid RNT medium. After the media had set, agar plugs were removed over the clear wells of the glass slide using a cork borer. The fungus was inoculated next to each well so that the hyphae grew into the wells and across the glass slide. After 2-3 days at 37°C, the agar was stripped off and the slides left to dry at RT. The mycelium was then fixed for 3 minutes in ethanol/chloroform/formalin (6 : 3 : 1). The slides were dipped in 95% methanol for 4 minutes and briefly washed with distilled water. After air drying, the slides could be stored at 4°C. For immunofluorescence detection, the fixed mycelium was covered with a droplet of antibody in PBS

and incubated at RT for 30 min in a moist chamber. The mycelium was given three 5 min washes with PBS. This was followed by an incubation with 20 µl Anti-rabbit-IgG (whole molecule) FITC conjugate (Sigma), diluted 1 : 40 in PBS for 30 minutes in the dark. After three washes with PBS, the coverslip was mounted using three 20 µl drops of Mowiol mountant. The slides could now be viewed immediately using a Zeiss Axiophot microscope (mercury lamp) or stored in the dark for a limited period. The nuclei could be viewed by DAPI fluorescence (345-375 nm) while any FITC fluorescence could be viewed independently using a different filter (450-490 nm).

2.24 GST fusion protein *in vitro* association assay

During the course of this study, the stringency of the GST fusion protein *in vitro* association assay (Kerr, Ransone et al. 1993) had to be increased in order to be able to observe specific protein-protein interactions (see chapter 6). The GST fusion protein (or the GST protein on its own for the negative control) was adsorbed to glutathione-S sepharose as described in section 2.15.5. The adsorbed protein was then washed twice with 900 µl PBS followed by a 3 minute incubation in 900 µl PBS/10 mg/ml BSA/5 mM DTT. After a 1 min centrifugation, the resin was resuspended in 100 µl of the same solution and 2 - 15 µl of *in vitro* translated protein were added. This mixture was incubated for 15 min at RT, followed by a 15 min incubation on ice. The resin was centrifuged, the supernatant removed for later analysis, and the resin washed with 900 µl PBS/10 mg/ml BSA/5 mM DTT followed by 3 washes with 900 µl PBS. The supernatant and the resin fraction were finally analysed by SDS-PAGE.

Chapter 3

Cloning of the *b* gene pair of the A6 mating type factor

Characterisation of the *C. cinereus* A42 mating type factor established (Kües, Richardson et al. 1992), that cross-hybridisation and transformation could be used to identify functionally identical alleles of specificity genes in different A factors. Cross-hybridisation correlated with the failure of a gene to promote clamp cell development in the transformation assay. A42 and A6 share the entire $A\alpha$ locus and also the *d1-1* gene in the $A\beta$ locus (see figure 3.1). A partial map of the A6 mating type factor was available at the start of this study but the region between *a2-1* and *d1-1* had not been identified and it was not known how many specificity genes were present in the A6 β locus. Figure 3.1 compares the map of the A42 factor (Kües, Richardson et al. 1992) with a partial map of the A6 factor (Richardson 1992) derived from fragments isolated from a plasmid genomic library constructed in the vector YRP12 (Yashar and Pukkila 1985).

The first aim of this project was to isolate the missing region of the A6 mating type factor. Since two of the specificity genes known to be present in A6 (*a2-1* and *d1-1*) were shared with A42, it was obvious that other genes must be present in the A6 β locus and that these would be responsible for the different mating specificities of A42 and A6. By isolating these other genes it would be possible to compare the DNA and predicted protein sequences of functionally different allelic forms of at least one pair of specificity genes and hopefully draw some conclusions as to how these genes/proteins interact on a molecular level.

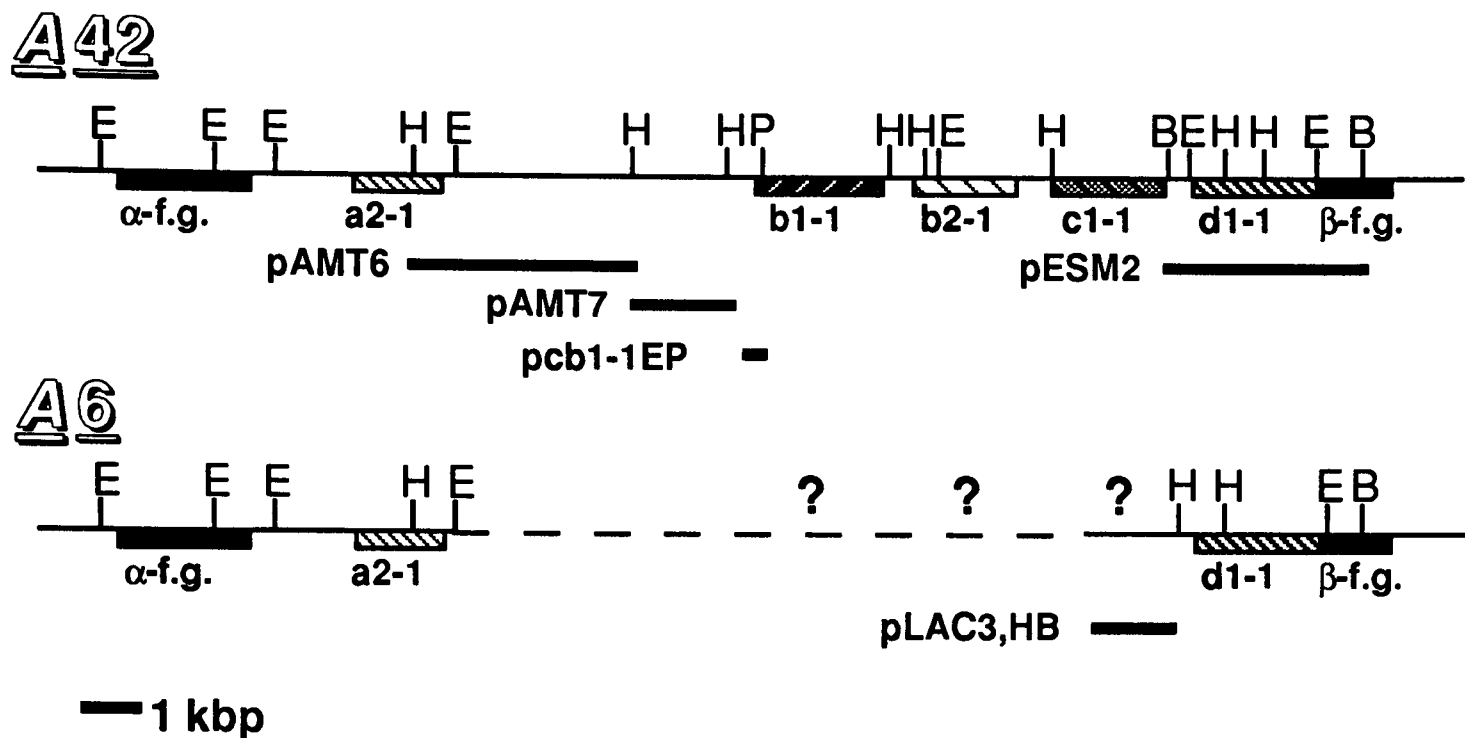


Figure 3.1: Organisation of the A42 factor and shared regions of the A6 factor. The dotted line indicates uncharacterised regions of A6. The fragments cloned in plasmids pAMT6, pAMT7, pcb1-1EP, pLAC3,HB are shown below each map and were used as hybridisation probes (E=*EcoRI*, H=*HindIII*, B=*BamHI*, P=*PstI*)

3.1 Screening a genomic library for genes of the A6 mating type factor

C. cinereus strain H9 is a wild type strain with A6B6 mating type. DNA from this strain had been used to construct a genomic library in the phage cloning vector λ GEM11 (Promega). A map of this vector is shown in figure 3.2. The central stuffer region was replaced in library construction by *MboI* digested genomic DNA fragments fused into the *BamHI* cloning site and with an average length of 15 kbp.

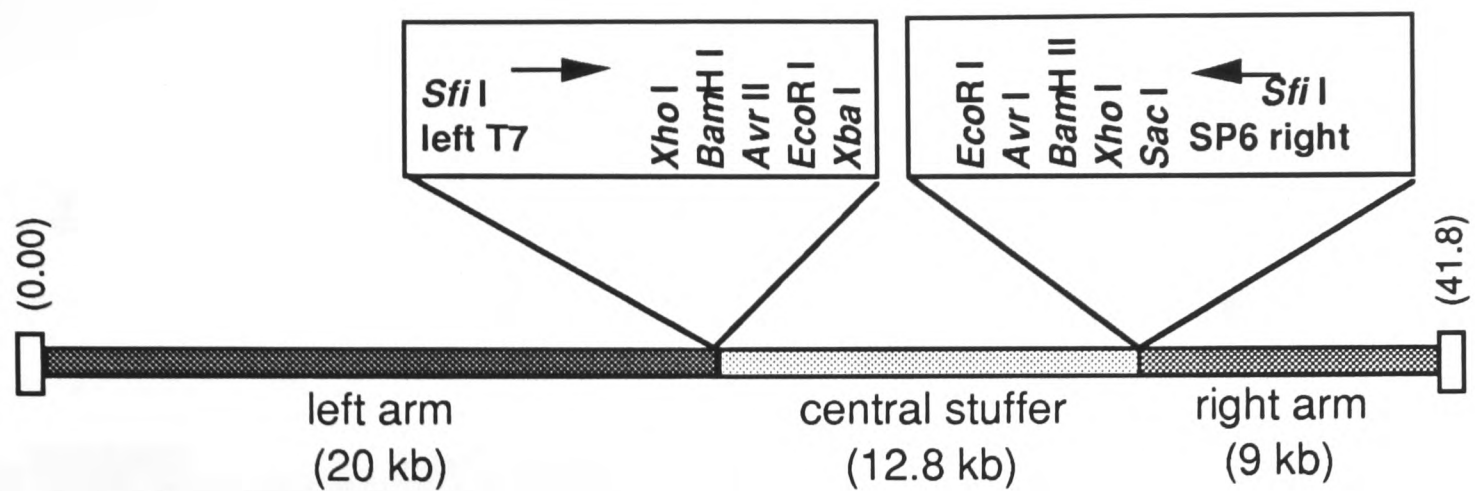


Figure 3.2: Map of the phage cloning vector λ GEM11, showing the two λ arms separated by the central stuffer region

In order to ensure a 99% probability of finding the entire *C. cinereus* genome in this library it was necessary to screen approximately 12,000 plaques. Two probes flanking the region of interest were used for this screen, a 1.9 kbp *Hind*III fragment from the homologous hole of A42 which is adjacent to *b1-1* (pAMT7) and a 1.9 kbp *Hind*III *Bam*HI fragment isolated from the YRP12 clone pLAC3 (pLAC3,HB) containing the sequence immediately 5' to the *d1-1* gene of A6 (see figure 3.1 for origin of hybridisation probes). Any clones hybridising to both should contain sequences spanning the entire missing region of the A6 β locus.

Ten positive plaques were identified in the initial library screen which hybridised to one or both probes. Up to 500 ng DNA isolated from each of these was subjected to single and double digestion with *Xho*I and *Hind*III (see figure 3.3(a)). Each clone contained an insert of approximately 15 kb. The orientation of the insert sequences was determined by Southern blot analysis using the two probes used to screen the library (pAMT7, pLAC3,HB) and pESM2 a 4 kbp fragment containing the A42 *d1-1* gene. The results of these hybridisations are shown in figure 3.3 (b),(c) and (d).

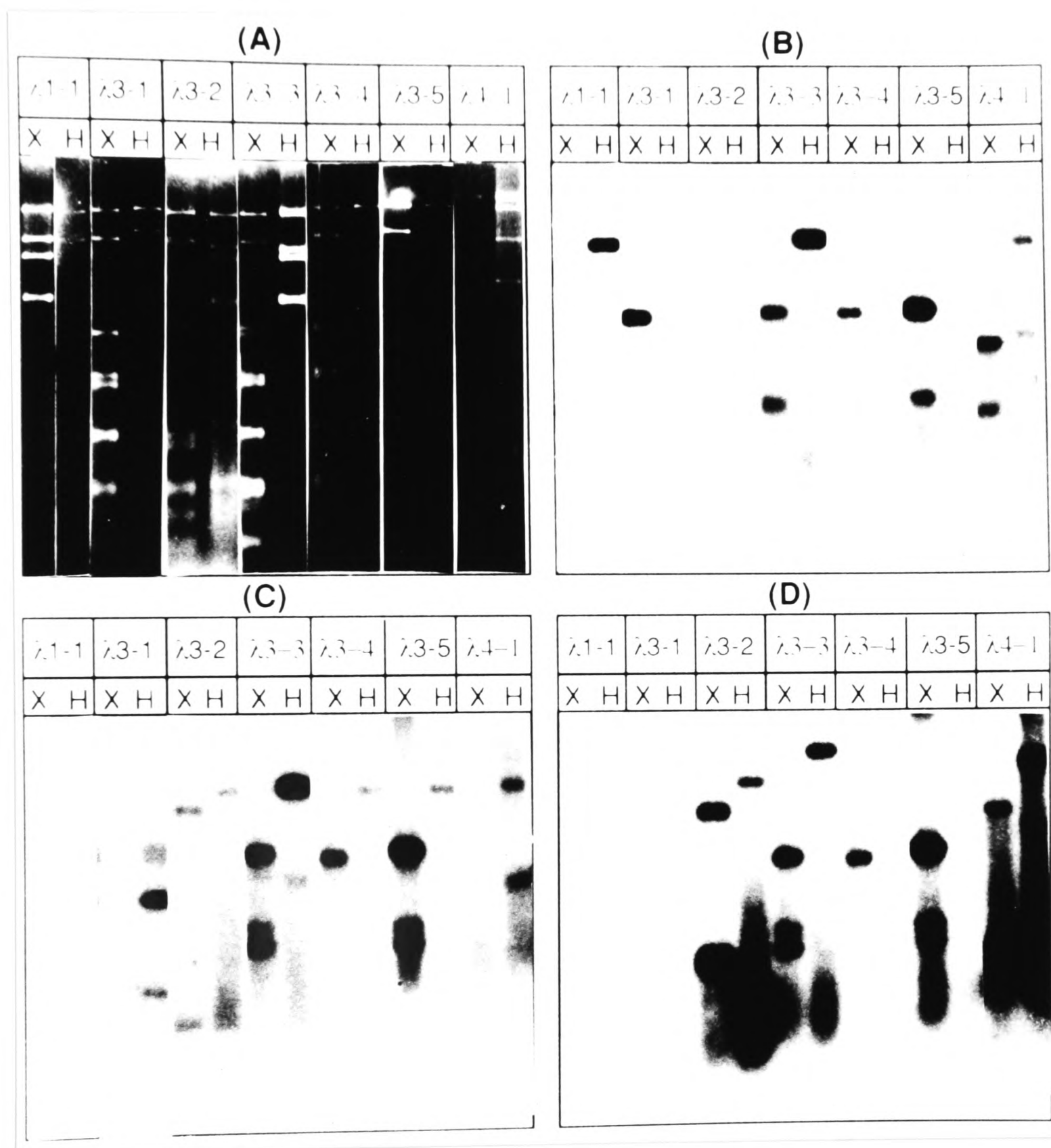


Figure 3.3: Southern blot analysis of phages λ 1.1, λ 3.1, λ 3.2, λ 3.4, λ 3.5 and λ 4.1 isolated from the H9 genomic library. Ethidium stained gel (A) with *Xho*I (X) and *Hind*III (H) restriction digests as indicated and three corresponding autoradiographs to show hybridisation to pAMT7 (B), pLAC3,HB (C) and pESM2 (D) (see figure 3.1 for origin of these probes). The sizes of hybridising fragments are given in table 3.1.

λ 1.1 and λ 3.5 were chosen for further analysis. The *Xho*I and *Hind*III digests of these two clones were reprobbed with pcb1-1EP (see figure 3.4), a 300 bp *Eco*R I/*Pst* I fragment containing the 3' end of the *b1-1* gene and pAMT6, a 4.2 kb *Hind*III fragment from the homologous hole of A42 (see figure 3.1). The results of all five hybridisations are summarised in Table 3.1. The 3' end of the A42 *b1-1* gene and the corresponding region of the A6 *b1* gene had previously been shown to be sufficiently homologous to cross-hybridise (Tymon, Kües et al. 1992). Thus the 3' fragment of *b1-1* contained in pcb1-1EP could be used to locate the position of the A6 *b1* gene.

Table 3.1: Identification of a fragment containing the *b* gene pair from the β locus of A6. The table lists the sizes of *Hind*III, *Xho*I and *Hind*III/*Xho*I fragments derived from the phages λ 1.1 and λ 3.5 hybridising to the five probes shown in figure 3.1 (gel isolated fragments were used to prepare all hybridisation probes; H=*Hind*III, X=*Xho*I)

	Probes used for hybridisation				
Phage/Digest	pAMT7	pcb1-1EP	pLAC3,HB	pAMT6	pESM2
λ 1.1/H	7.5 kb	7.5 kb	-	4.2 kb	-
λ 3.5/H	7.5 kb	7.5 kb	7.5 kb	4.2 kb	24.0 kb
λ 3.5/H-X	1.7 kb	2.5/0.6 kb	2.4 kb	1.7/1/0.8/0.5 kb	1.4 kb

These hybridisation data together with the fragment sizes determined from the agarose gel allowed the construction of preliminary maps of λ 1.1 and λ 3.5 based on *Hind*III restriction sites (figure 3.4). The sequence cloned in λ 1.1 hybridised to probes from the homologous hole and the homologous 3' end of the *b1-1* gene but not to sequences from the region adjacent to the *d1-1* gene. All five probes hybridised to fragments derived from lambda clone λ 3.5 indicating that it contained the entire region of interest extending from the homologous hole to the *d1-1* gene. The hybridisation data also showed

that this unique A6 sequence was contained within a 7.5 kb *Hind*III fragment (figures 3.3 and 3.4).

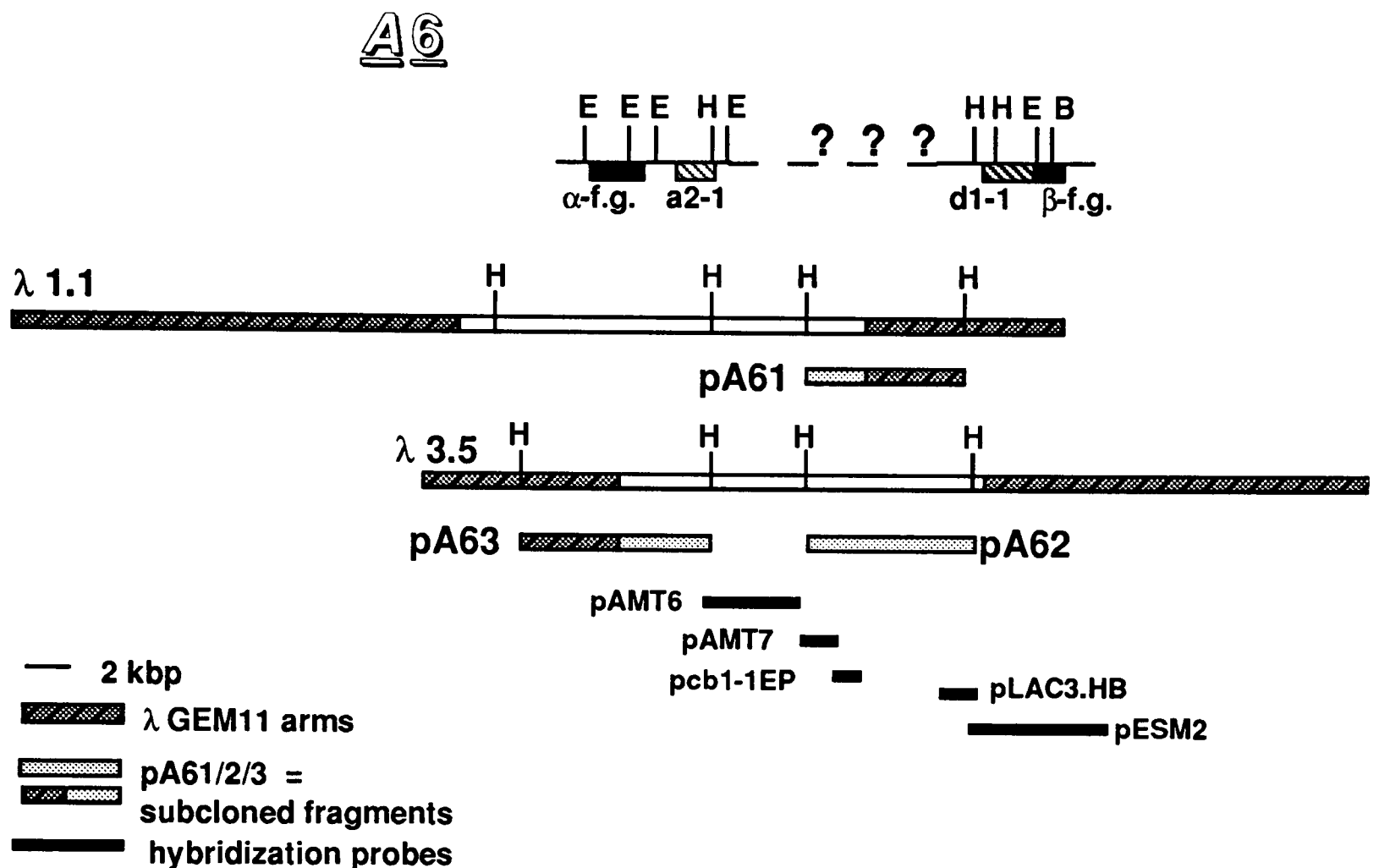


Figure 3.4: Map of overlapping λ phages 1.1 and 3.5. A preliminary restriction map of the A6 factor (see figure 3.1) is included for reference (for origin of the hybridisation probes pAMT6, pAMT7, pcb1-1EP, pESM2 and pLAC3,HB see figure 3.1). Also shown are DNA fragments that were subcloned from λ 1.1 and λ 3.5 into pBluescript KS- to yield pA61, pA62 and pA63 (H=*Hind*III, E=*Eco*RI, B=*Bam*HI).

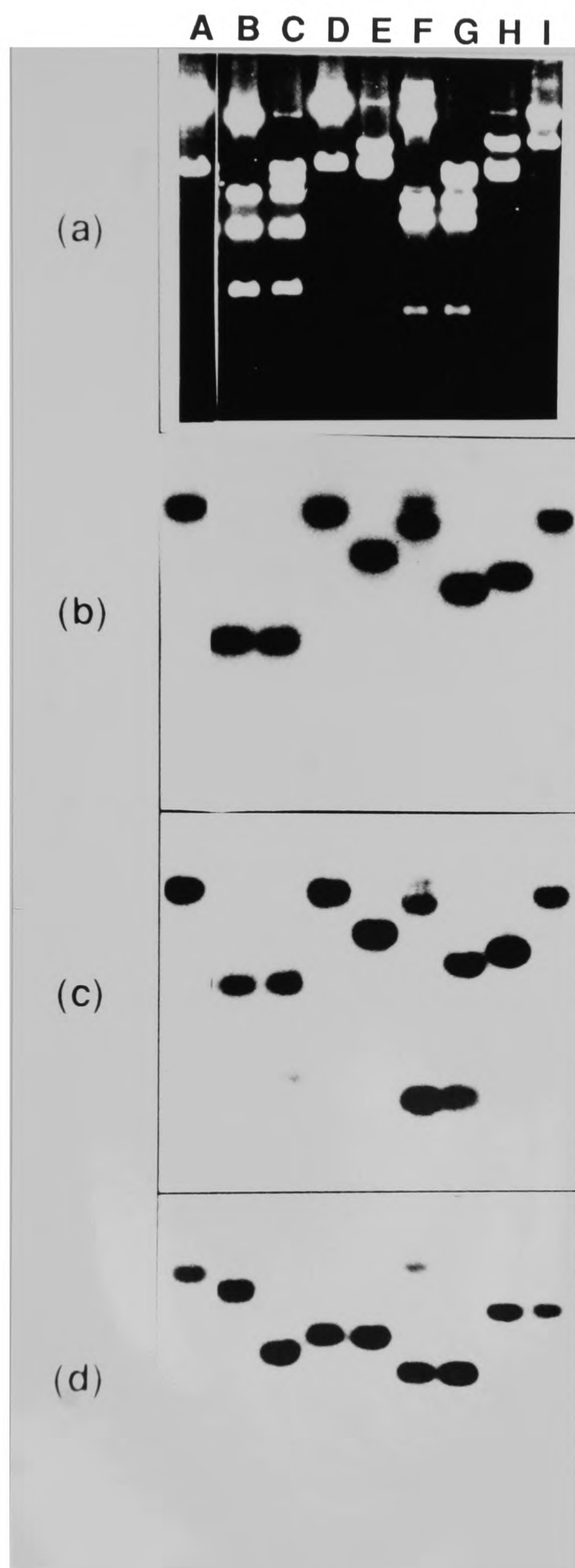
3.2 Subcloning of the A6 mating type locus genomic DNA fragments in pBluescript II, KS-.

Three *Hind*III fragments were subcloned into the vector pBluescript II KS-. The 7.5 kb *Hind*III DNA fragment in pA61 is from λ 1.1 and contains part of the homologous hole and the 3' end of the A6 *b1* gene and 4.5 kb of DNA

that originates from the shorter arm of the λ GEM11 cloning vector (figure 3.4). The 7.5 kb *Hind*III fragment in pA62 from λ 3.5 contains part of the homologous hole adjacent to the A6 *b1* gene, the complete *b* gene-pair of the A6 mating type locus and possible additional specificity genes. pA63 contains an 8.5 kbp *Hind*III fragment from λ 3.5, which comprises part of the A6 α locus including the *a2-1* gene, and approximately 4.5 kbp of DNA sequence from the shorter arm of λ GEM11 (figure 3.4). All subsequent experiments concentrated on clone pA62 which contained the A6 *b* gene-pair.

As described in the introduction, a compatible A gene interaction should only occur between different alleles of the same gene-pair. From the transformation studies that identified the A42 β locus, it can be seen that the A6 factor must contain a functional pair of *b* genes because both *b1-1* and *b2-1* from A42 independently elicit the formation of clamp cells when transformed into an A6 host (Kües, Richardson et al. 1992). The predicted A6 *b* genes were designated *b1-3* and *b2-3* to distinguish them from the allele pairs present in A42 (*b1-1*, *b2-1* (Kües, Richardson et al. 1992)) and in another cloned A factor, A43 (*b1-2*, *b2-2* (May, Le Chevanton et al. 1991)).

Single and double digests of pA62 with *Xho*I, *Hind*III, *Eco*RI, *Pst*I and *Bam*HI were analysed by Southern blotting using pAMT7, pcb1-1EP and pLAC3,HB as hybridisation probes (Figure 3.5). pAMT7 identified the homologous hole, probe pcb1-1EP the 3' end of the *b1* gene and probe pLAC3,HB the 5' region of the *d* gene. The origin and position of these hybridisation probes is shown in figure 3.1 and the results of this analysis are summarised in Table 3.2.



restriction digests of pA62:

A: *Hind*III
 B: *Hind*III/*Xho*I
 C: *Xho*I
 D: *Eco*RI
 E: *Hind*III/*Eco*RI
 F: *Pst*I
 G: *Hind*III/*Pst*I
 H: *Hind*III/*Bam*HI
 I: *Bam*HI

Figure 3.5: Southern blot analysis of pA62. (a) ethidium stained gel with restriction digests as indicated and three corresponding autoradiographs obtained using (b) pAMT7, (c) pcb1-1EP, (d) pLAC3,HB (see figure 3.1 for origin of these probes). The sizes of hybridising fragments are given in table 3.2.

Table 3.2: Results of Southern analyses used to construct a restriction map of pA62. The table shows the sizes of fragments to the probes (origin and position of probes is shown in figure 3.1)

Restriction enzyme digests	Probes used for Southern Analyses		
	pAMT7	pcb1-1EP	pLAC3,HB
<i>Hind</i> III	7.5 kb	7.5 kb	7.5 kb
<i>Hind</i> III/ <i>Xho</i> I	1.7 kb	2.2/0.7 kb	2.5 kb
<i>Xho</i> I	1.7 kb	2.2/0.7 kb	5.5 kb
<i>Eco</i> RI	7.0 kb	7.0 kb	3.2 kb
<i>Hind</i> III/ <i>Eco</i> RI	4.0 kb	4.0 kb	3.2 kb
<i>Pst</i> I	5.3 kb	5.3/0.7 kb	2.2 kb
<i>Hind</i> III/ <i>Pst</i> I	2.3 kb	2.3/0.7 kb	2.2 kb
<i>Hind</i> III/ <i>Bam</i> HI	3.2 kb	3.2 kb	4.2 kb
<i>Bam</i> HI	6.2 kb	6.2 kb	4.2 kb

It was now possible to construct the more detailed restriction map of pA62 illustrated in Figure 3.6. The Southern blot shown in Figure 3.5 was also probed with the *A42 d1-1* gene present in pESM2 and as expected, this probe did not hybridise, confirming that the *d1-1* gene was not present in pA62. This Southern blot was also probed with a sequence from the homologous hole in pAMT6, which also did not hybridise, confirming that pA62 did not contain sequences beyond those defined by pAMT7.

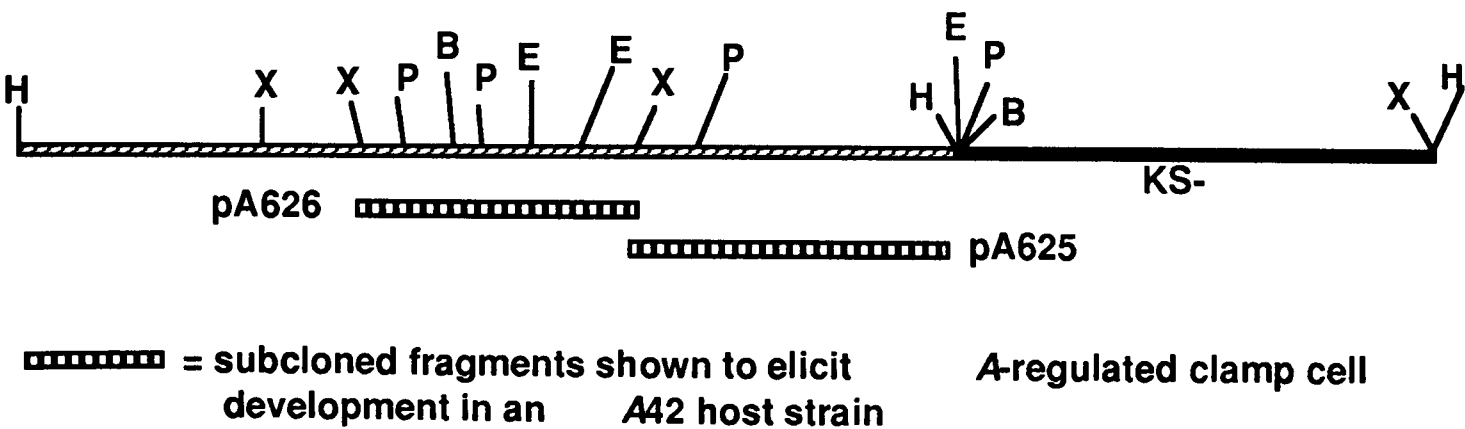


Figure 3.6: Restriction enzyme map of pA62 (B=*Bam*HI, E=*Eco*RI, H=*Hind*III, P=*Pst*I, X=*Xho*I, KS-= pBluescript KS- vector sequence)

Hybridisation with pcb1-1EP located the 3' end of the A6 *b1* gene on a 0.8 kb *XhoI* fragment. The 7.5 kb *HindIII* fragment thus contains the entire A6 *b1* gene and is large enough to contain the a *b2* gene.

The 7.5 kbp fragment contained in pA62 is unlikely to contain more than two genes. 2 kbp are occupied by DNA originating from the non-coding homologous hole leaving just 5.5 kbp of DNA to code for the specificity genes. The *b* gene-pair in A42 occupies approximately 5 kbp of DNA sequence and since the 5' end of the A6 *d* gene is directly adjacent to the *HindIII* cloning site in pA62, there would be less than 1 kbp of DNA sequence left for any additional A6 genes.

3 3 Identification of the A6 *b* genes by transformation

Two fragments of pA62 were subcloned and used to transform the *C. cinereus* strain LN118 (A42B42). pA625 was obtained by religating a 5.5 kbp fragment derived from *XhoI* digestion of pA62, and pA626 by ligating the 2.2 kbp *XhoI* fragment into *XhoI* digested pBluescriptII KS-. The sequences present in these two subclones are indicated in figure 3.6.

Both clones promoted the formation of unfused clamp cells following transformation into the A42 host. The hybridisation data described in table 3.2 showed that the homologous 3' end of the A42 *b1-1* gene hybridised to a 0.7kbp and a 2.2 kbp band in *XhoI* digested pA62. It was therefore concluded that pA626 (a subclone of this 2.2 kbp *XhoI* fragment) contained the A6 *b1-3* gene. The adjacent clone, pA625, should therefore contain the *b2-3* gene.

The transformation data confirmed that both *b* genes from the A6 mating type factor had been cloned. The first objective of the thesis was thereby achieved and it was now necessary to sequence these genes. It was essential at this stage to construct a cDNA library from an A6 monokaryon since the interpretation of sequence data is greatly facilitated by obtaining cDNA sequence and cDNA clones would be required at a later stage for the preparation of expression constructs to characterise A proteins on a molecular level.

3.4 Preparation of a cDNA library of *C. cinereus* strain H9

Poly A(+) RNA was isolated from *C. cinereus* strain H9 (A6B6) following the procedure outlined in chapter 2. The good quality of the RNA was proved by the sharp bands obtained in Northern blot analysis. The probe used was the 2.5 kb *HindIII/XhoI* insert of the clone pA625 which identified the b2-3 message (figure 3.7).

poly A(+) RNA was used to prepare a cDNA library using the λ ZAP cloning vector (Stratagene) with some modifications to the supplier's recommended procedure (see chapter 2). The yield of cDNA was 2 μ g, calculated from the results of Cerenkov counting. The products of the 1st and 2nd strand cDNA synthesis reactions were analysed by alkaline agarose gel electrophoresis. An autoradiograph of this gel (figure 3.8) showed that cDNA of up to 2.3 kbp in length had been produced.

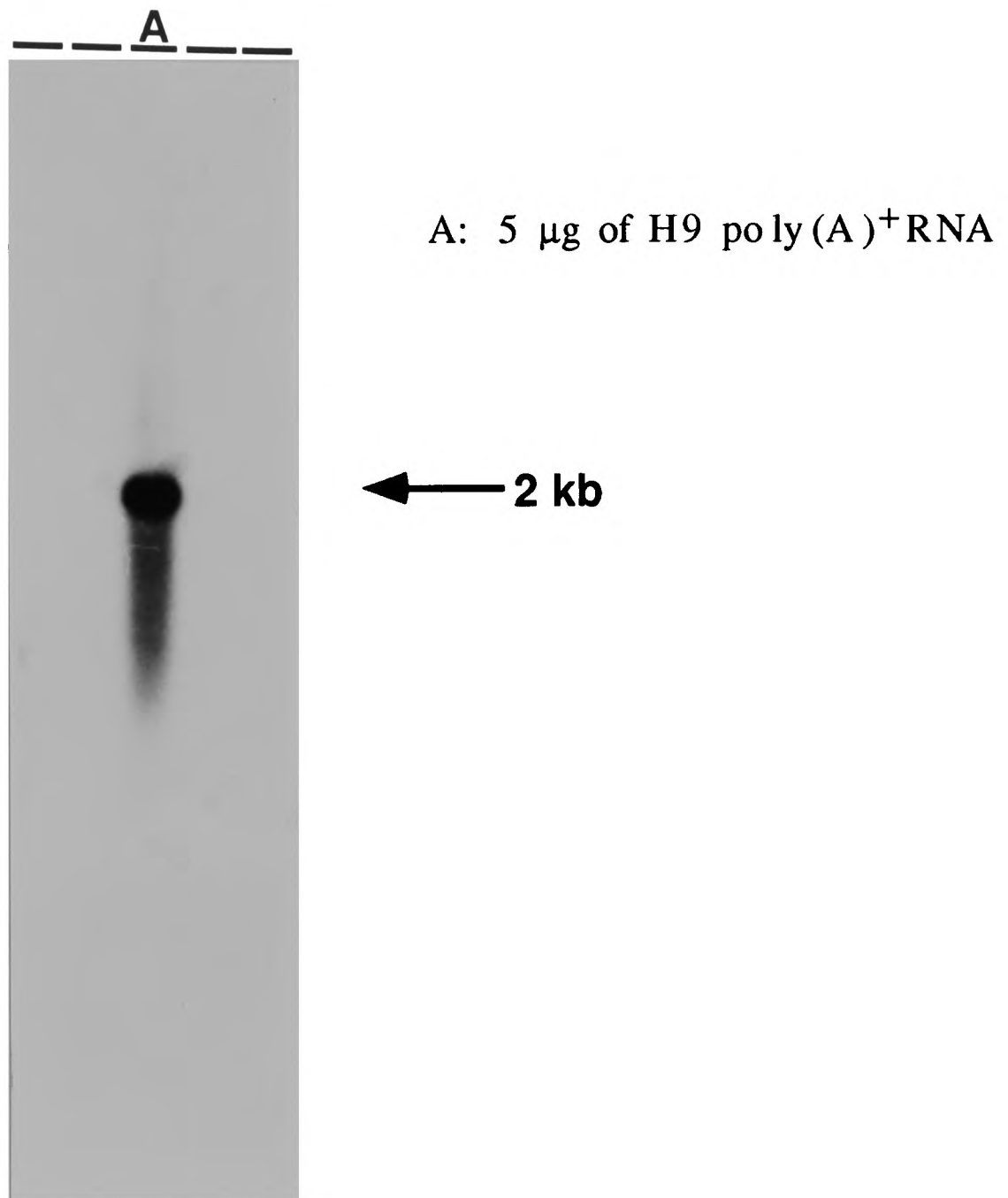


Figure 3.7: Northern blot analysis of H9 poly(A)⁺ RNA using a XhoI/HindIII fragment from pA625 (see figure 3.6) as hybridisation probe to identify the mRNA of the *b2-3* gene (message size estimated by comparison to 18S and 28S rRNA bands and GIBCO-BRL RNA ladder, both visualised by ethidium staining).

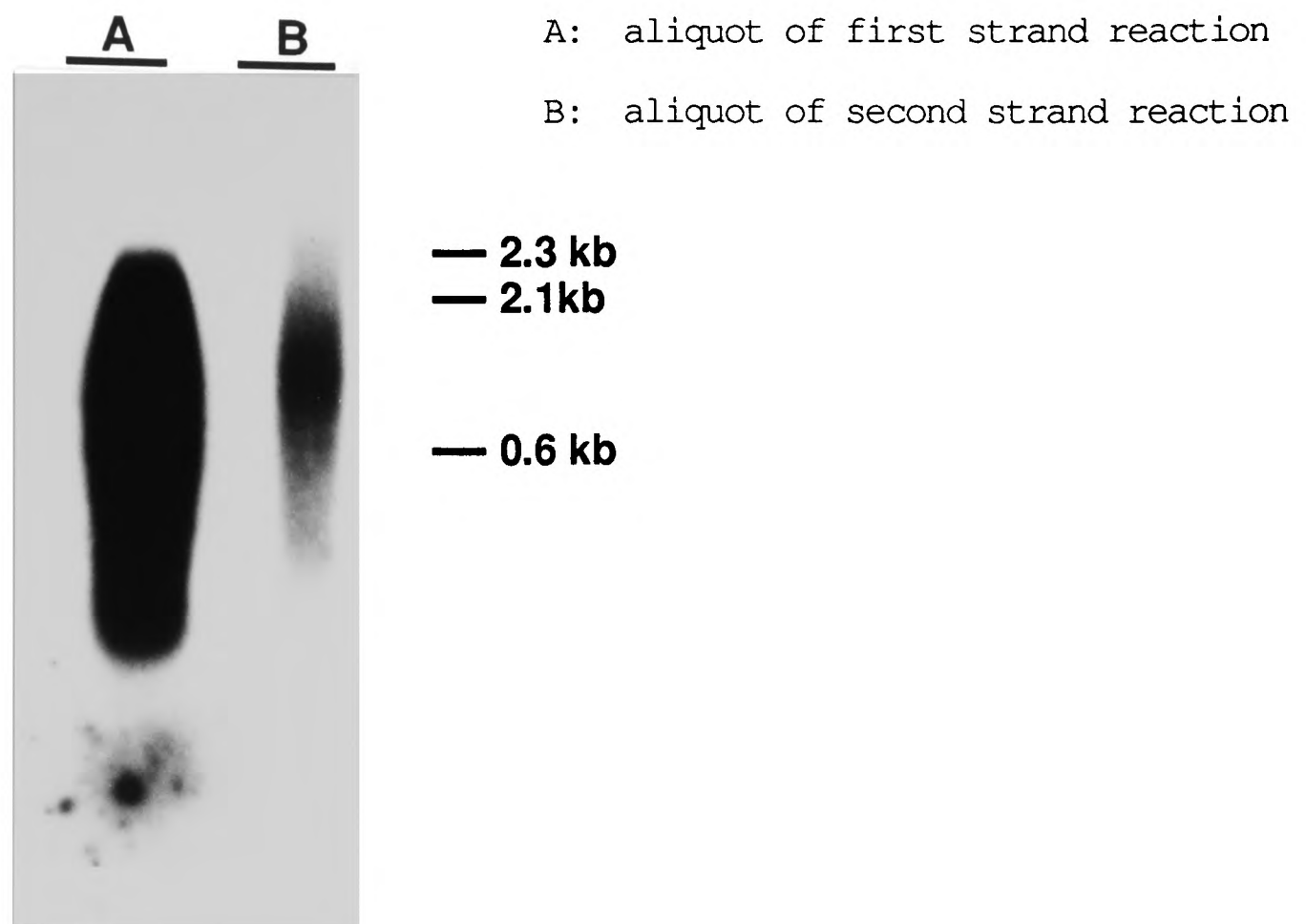


Figure 3.8: Analysis of first and second strand cDNA synthesis by alkaline agarose gel electrophoresis (method described in chapter 2).

The cDNA was cloned into the phage vector λ ZAP II according to the protocol outlined in the kit manual (Stratagene). After the size separation step (spin column), the amount of cDNA in the first two fractions was calculated to be 80 ng. The first five fractions were analysed by gel electrophoresis. Figure 3.9 shows that fractions 1 and 2 contained cDNA in the size range 0.5 to 3 kb and these were used for the construction of the cDNA library. The primary library contained 4,400,000 pfu with only one in 400 non-recombinant phages. The library was amplified once and a total of 6×10^{11} pfu were obtained. The average insert size was determined using the two approaches described in chapter 2 and found to be 1.1 kbp.

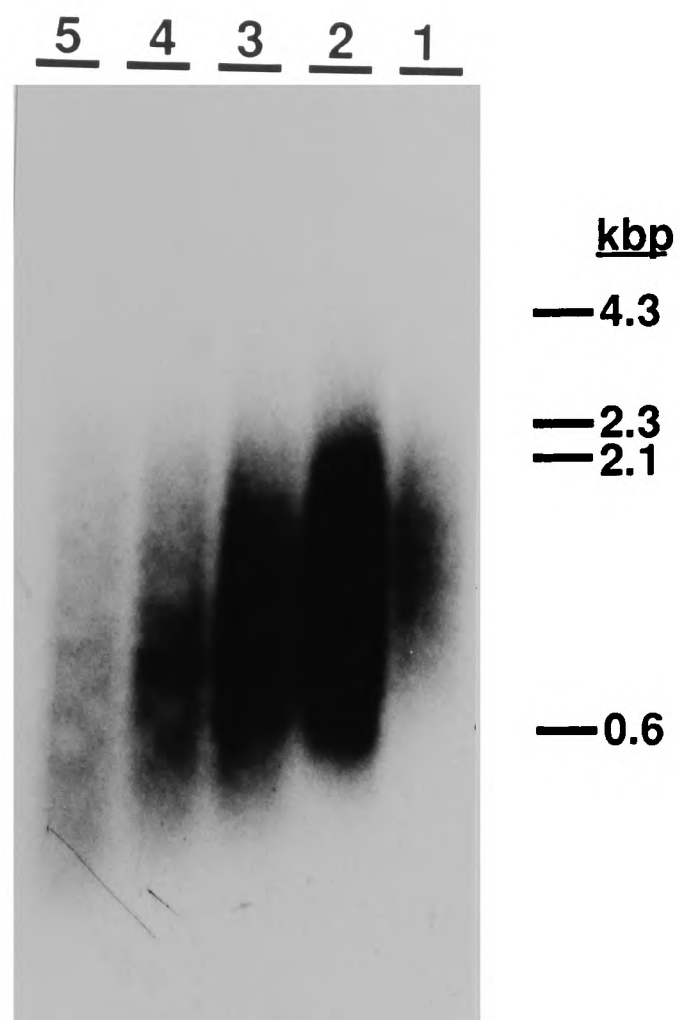


Figure 3.9: Analysis of size separated cDNA by agarose gel electrophoresis. Track numbers correspond to the fraction numbers of the size separated cDNA.

3.5 Screening of the H9 cDNA library for clones of the A6 *b* genes

90,000 pfu of the amplified λ ZAP library were screened for the *b* gene-pair from the A6 β locus using as probe a 4 kbp *Bam*HI/*Hind*III band containing part of *b1-3* and the entire *b2-3* gene (A6219). Two positive plaques were identified. The library was screened again with a smaller 2 kbp *Pst*I fragment containing the 5' ends of *b1-3* and *b2-3* (A6211). This second screen gave two additional positive plaques. Second round screening allowed the isolation of single plaques hybridising to the probes.

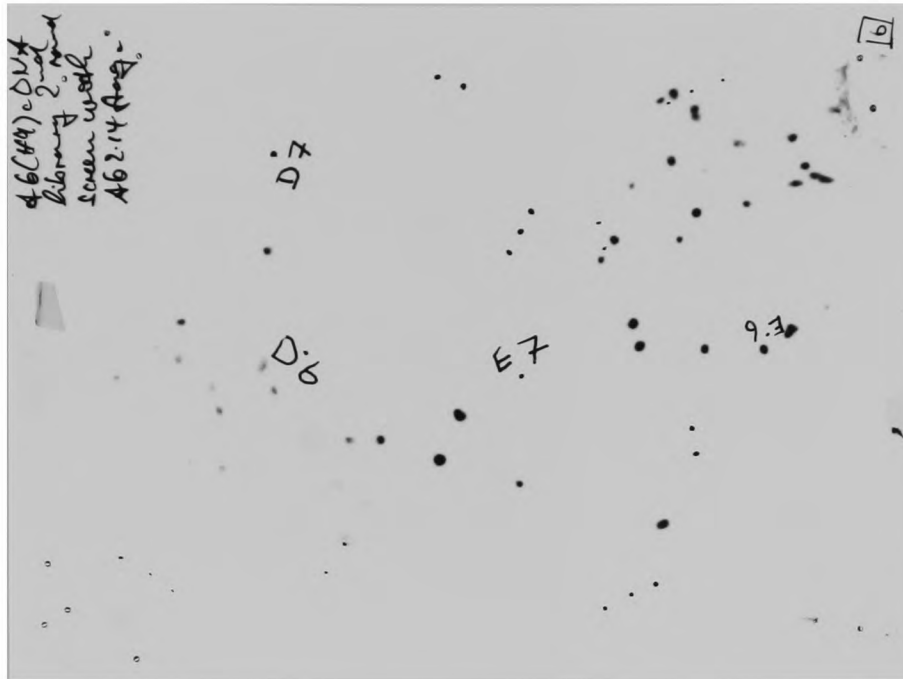
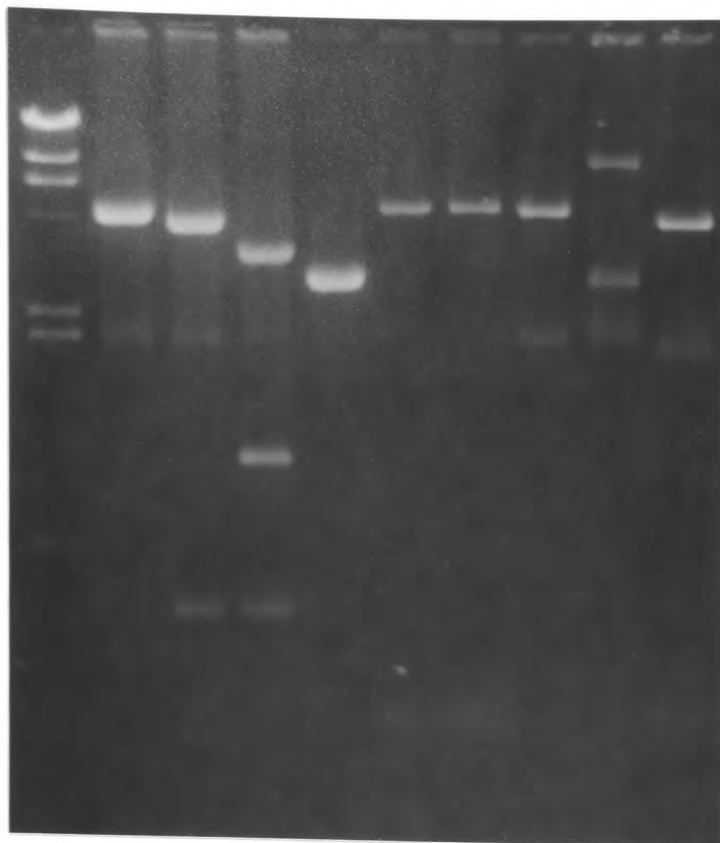


Figure 3.10: Second round screen of the A6 (H9) cDNA library using a 4 kbp *Bam*HI/*Hind*III fragment containing *b1-3* and *b2-3* sequences as hybridisation probe (A6214)

pBluescript clones were released from the λ ZAP phages. DNA minipreps were prepared for all positive cDNA clones and figure 3.11 illustrates restriction digests of these clones. DNA sequencing (chapter 4) was used to identify which of the two A6 *b* genes each clone corresponded to. One cDNA corresponded to the *b1-3* gene but was found to be a hybrid with 500 bp of an unidentified message cloned behind 1.3 kbp of *b1-3* cDNA sequence. The remaining three cDNA clones originated from the *b2-3* mRNA. One was a full length cDNA (1.95 kbp, pb2-3cDNA2), one contained a nearly full length insert (1.87 kbp, pb2-3cDNA1) and the third clone (pb2-3cDNAh) was a hybrid clone containing a 1.8 kb insert which was derived partly from *b2-3* cDNA cloned and partly derived from a gene coding for a ribosomal protein (see appendix for details). These results suggest that the A6 cDNA library is still representative of the primary library after one round of amplification since all four positive clones originated from different mRNAs.

A B C D E F G H I J



- A: λ HindIII size marker
- B: pb1-3cDNA, *Eco*RI
- C: pb1-3cDNA, *Bam*HI
- D: pb1-3cDNA, *Bam*HI/*Eco*RI
- E: pBluescript SK-, *Eco*RI
- F: pb2-3cDNA2, *Bam*HI/*Eco*RI
- G: pb2-3cDNA2, *Eco*RI
- H: pb2-3cDNA1, *Bam*HI
- I: pb2-3cDNA1, *Eco*RI
- J: pb2-3cDNAh, *Eco*RI

Figure 3.11: Ethidium stained agarose gel to show restriction digests of cloned cDNAs.

Figure 3.11 shows that the pb1-3cDNA clone has lost the 5' *Eco*RI cloning site. *Eco*RI linearises this clone but the single *Eco*RI site is not next to the *Bam*HI site in the plasmid linker. Otherwise the *Bam*HI and the *Bam*HI/*Eco*RI double digest should produce an identical restriction pattern. The single *Eco*RI site in this clone must be in the unidentified part of the hybrid insert. The figure also shows that the *Eco*RI cloning site was absent in the pb2-3cDNA1 clone. These observations were later confirmed by DNA sequencing (see chapter 4).

Chapter 4

DNA sequence analysis of the *b1-3* and *b2-3* genes and construction of chimeric *b* genes

The *b1-1* and *b2-1* genes from A42 have been shown to be divergently transcribed, to have message sizes of 2.5 and 2.1 kb and to code for proteins of 632 and 523 amino acids respectively (Tymon, Kües et al. 1992) (Kües, Richardson et al. 1992)(+ unpublished). The distinctive features of these two proteins are dissimilar homeodomains which have been defined as HD1 (present in *b1-1*) and HD2 (present in *b2-1*) (Kües, Richardson et al. 1992). pA62 contains the corresponding *b* gene alleles from the A6 factor, *b1-3* and *b2-3*. Apart from the 3' ends of the *b1* genes, the A42 and A6 genes do not cross hybridise in Southern analysis but their function in regulating sexual development is absolutely conserved. By sequencing the A6 *b* genes, it was hoped to identify similar regions in allelic proteins that might suggest conserved functions. At the same time, identification of dissimilar regions would indicate features that determine allele specificity.

4.1 Sequencing of the A6 *b1-3* and *b2-3* genes

Southern hybridisation experiments using probes from the β locus of A42 and transformation studies showed that the A6 *b1-3* and *b2-3* gene should be contained within a 5.5 kbp region of pA62. This DNA sequence, extending from the first *Xho*I site in the homologous hole to the *Hind*III cloning site near the 5' end of *d1-1* was determined as indicated in figure 4.1. The sequence is presented in the appendix and aligned to the corresponding sequence from the A42 *b* genes.

Figure 4.1 summarises the strategy used to obtain the DNA sequences of *b1-3* and *b2-3*. The genes were sequenced in both directions on genomic DNA template. In addition, the cDNA sequence for most of the coding region was also determined.

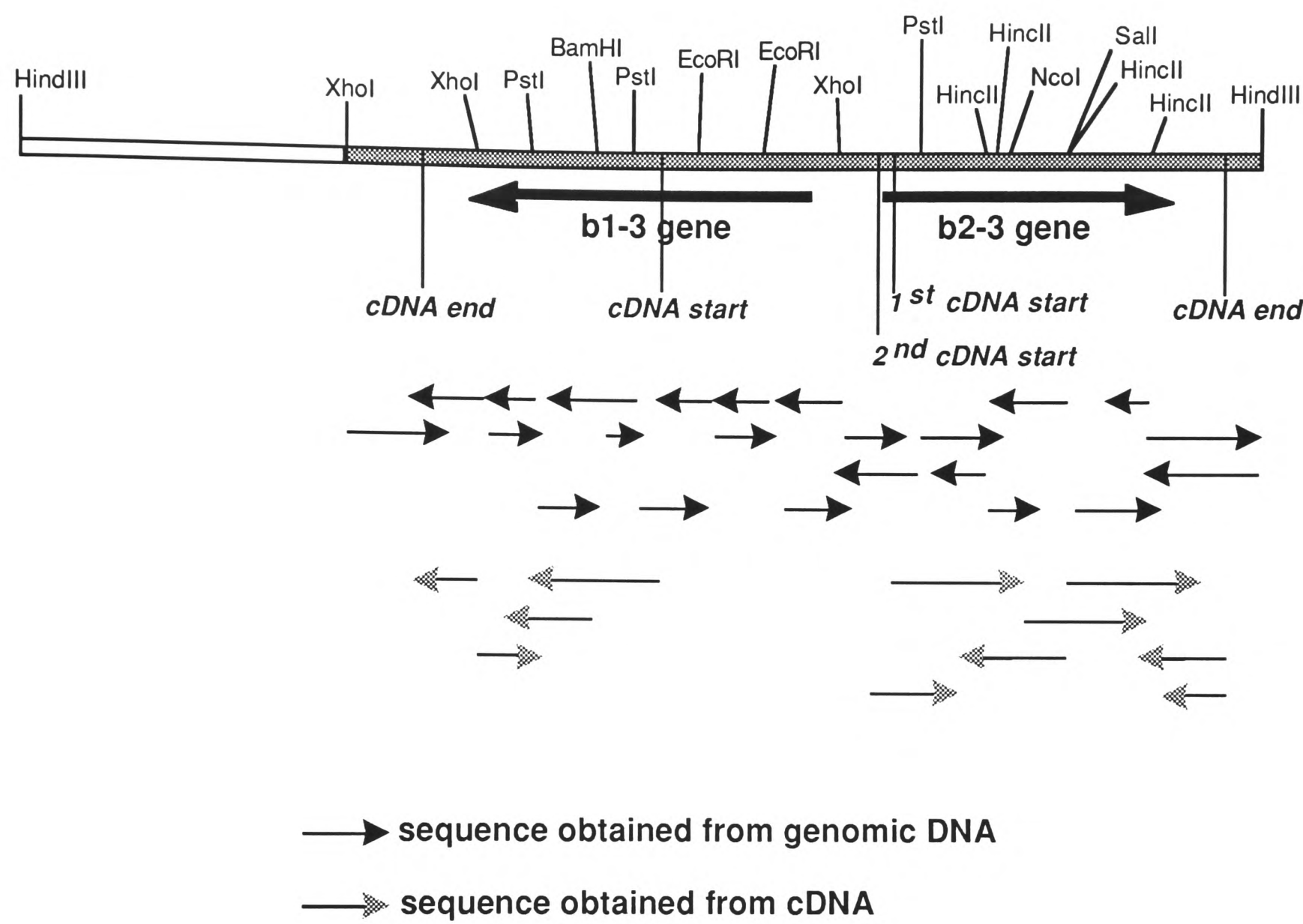


Figure 4.1: Strategy employed for sequencing the A6 *b* genes

Tables 4.1 and 4.2 list the subclones used for sequencing *b1-3* and *b2-3*. Numbering starts with 1 at the *Xho*I site within the homologous hole adjacent to the 3' end of *b1-3* and ends with 5572 at the *Hind*III site beyond *b2-3*. The 1.7 kbp DNA sequence between the *Hind*III site and the first *Xho*I site in pA62 was not determined since earlier studies had shown that the homologous hole is a non-coding region (Kües, Richardson et al. 1992).

Table 4.1: Subclones generated for sequencing the *b1-3* and *b2-3* genomic DNA (all subclones in pBluescript KS+, KS-SK+ or SK-; +/- indicates whether or not a clone had been sequenced with the respective primer; s. s. indicates sequencing with single stranded DNA template, ^a indicates that a subclone started approximately 1.7 kbp outside of the region sequenced)

name	cloning site	vector	start	end	forward	reverse
pA622	<i>EcoRI/HindIII</i>	KS-	^a -1700	2153	+	-
pA624	<i>EcoRI</i>	KS-	2147	2556	+	+ s. s.
pA625	<i>XhoI/HindIII</i>	KS-	3002	5572	+	+
pA626	<i>XhoI</i>	KS-	799	3007	+	+ s. s.
pA628	<i>XhoI</i>	KS-	1	804	+	+ s. s.
pA629	<i>HindIII/PstI</i>	KS-	^a -1700	1129	+	-
pA6210	<i>PstI</i>	KS-	3494	5572	-	+ s. s.
pA6211	<i>PstI</i>	KS-	1749	3499	+	+ s. s.
pA6212	<i>PstI</i>	KS-	1124	1755	+	+ s. s.
pA6214	<i>BamHI/HindIII</i>	KS+	1521	5572	+ s. s.	-
pA6216	<i>SalI/HindIII</i>	KS-	4395	5572	-	+ s. s.
pA6217	<i>PstI/SalI</i>	KS+	3494	4400	+ s. s.	-
pA6218	<i>HincII/HindIII</i>	KS-	4907	5572	-	+ s. s.
pA6219	<i>HincII</i>	KS+	4395	4912	-	+
pA6221	<i>HincII</i>	KS+	3956	4400	-	+
pA6222	<i>EcoRI/XhoI</i>	KS-	2550	3007	+	-
pA6223	<i>PstI/HincII</i>	KS+	3494	3911	-	+

Table 4.2: Subclones used for sequencing of the A6 *b* gene pair cDNA clones (legend as for table 4.1). Mutated *EcoRI* sites are underlined. The pb1-3 Δ *Sal* clone was a deletion of an internal *SalI* fragment of the pb1-3cDNA clone.

name	cloning site	vector	start	end	forward	reverse
pb2-3cDNA1	<u>EcoRI</u> / <i>XhoI</i>	SK-	3334	5349	+	+ s. s.
pb2-3Nco	blunt/ <i>XhoI</i>	SK-	4038	5349	-	+ s. s.
pb2-3PSal	<i>SalI</i> / <i>PstI</i>	SK+	3494	4400	+ s. s.	-
pb2-3Sal3'	blunt/ <i>XhoI</i>	SK-	4395	5349	-	+ s. s.
pb2-3cDNA2	<u>EcoRI</u> / <i>XhoI</i>	SK-	3257	5356	+	+ s. s.
pb1-3cDNA	<u>EcoRI</u> / <i>XhoI</i>	SK-	496	1871	+	+ s. s.
pb1-3Bam	<i>Bam</i> HI/ <i>XhoI</i>	SK-	496	1526	-	+
pb1-3Xho	<u>EcoRI</u> / <i>XhoI</i>	SK-	799	1871	+	-
pb1-3 Δ Sal	<u>EcoRI</u> / <i>XhoI</i>	SK-	496	856	-	+

As noted in the previous chapter, the *EcoRI* cloning site of the λ ZAP cloning vector was destroyed in pb1-3cDNA and pb2-3cDNA1. This was apparent from restriction digestion (see figure 3.11) and confirmed by DNA sequencing as shown in figure 4.2.

The sequences of *b1-3* and *b2-3* were aligned to the corresponding A42 *b1-1* and *b2-1* alleles (see appendix). Overall the two sequences show 63% identity, but there is a clear region of higher similarity at the 3' end of the *b1* genes and in the flanking part of the homologous hole (82% for the first 1.5 kbp sequenced) compared with the major coding region of the two *b* genes (57% for the final 4 kbp). Figure 4.4 shows a plot of the overall similarity derived from this comparison. This plot was obtained using the plotsimilarity program of the GCG DNA sequence analysis package using a window size of

250 bp. The high level of similarity seen in the homologous hole decreases within the *b* gene-pair coding region. Within the coding region there are two peaks of relatively high homology which correspond to the homeodomain coding sequences. These would be expected to be highly conserved in order to allow the different mating type proteins to function equally well as sequence specific DNA binding proteins. The sequence lying between the *b1* and *b2* genes is the least conserved. This should effectively prevent recombination between the two *b* gene-pairs and will have the effect of keeping particular pairs of alleles together during evolution. This is essential to maintain self-incompatibility. If recombination could occur between *b1* and *b2* genes, compatible *b1* and *b2* alleles would be brought together in the same monokaryon and could promote A regulated development without the need to mate.

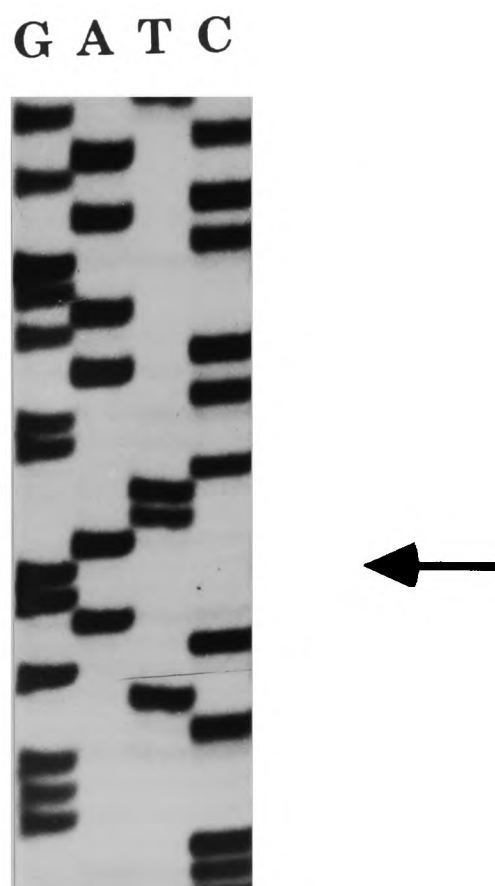


Figure 4.2: Part of an autoradiograph of a sequencing gel showing the sequence at the 5' cloning site of pb2-3cDNA1. The arrow indicates the position of the missing A in the GAATTC *Eco*RI recognition site.

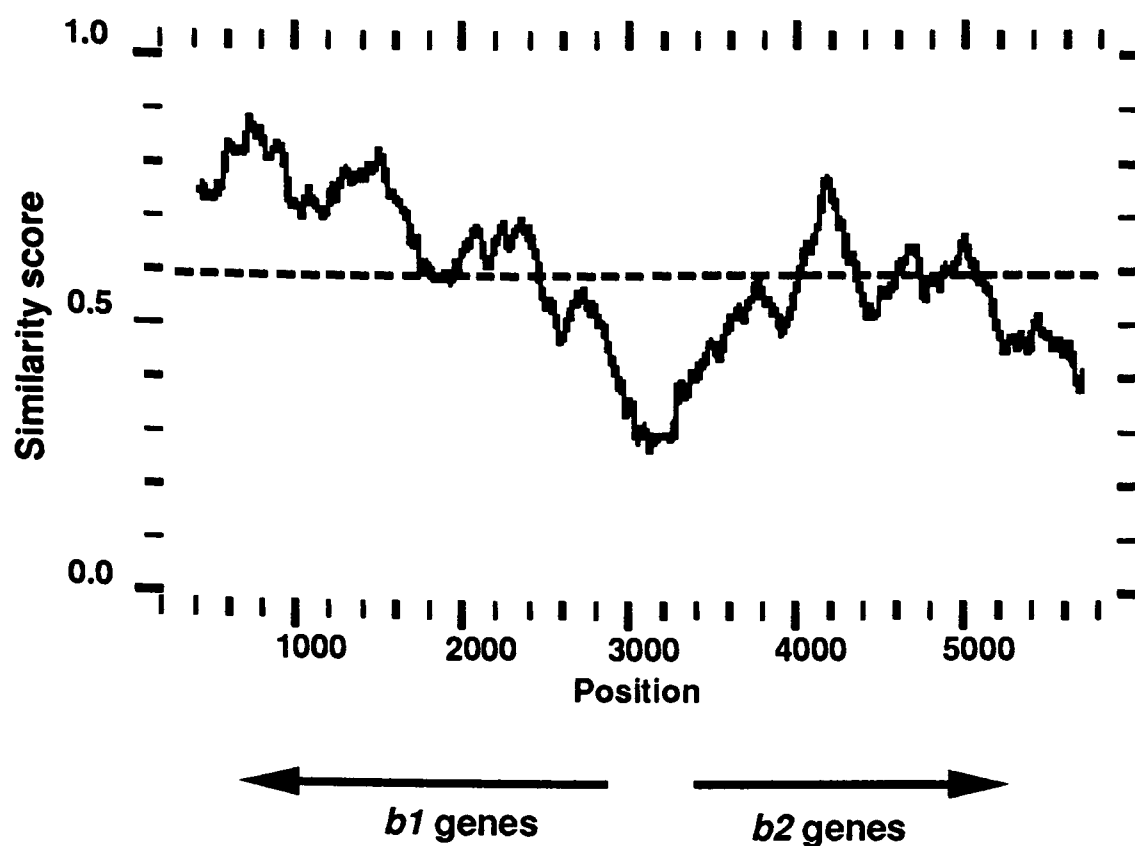


Figure 4.4: Comparison of the *b* gene-pair DNA sequences of the A6 and A42 mating type factors. Points used to generate the graph represent the percentage identity calculated over 250 bp regions. The position of the *b1* and *b2* genes is included for reference.

4.2 Gene structure as revealed by primary sequence data

The DNA sequence of the A6 *b* gene-pair shows that there are two divergently transcribed genes like the equivalent *b* gene-pair from the A42 factor (Kües, Richardson et al. 1992). The complete DNA sequences are shown in figures 4.4 and 4.5. Sequence assumed to code for the *b* proteins is printed in capital letters, the predicted start and stop codons are printed in bold. The translation of the coding sequence is shown in one letter code above the DNA sequence.

The sequence of the genomic DNA fragment cloned in pA625 is shown in figure 4.5. Figure 4.4 shows the reverse complement of the sequences present in both pA626 and pA627. It is evident that 16 codons have been deleted from the end of the *b1-3* gene cloned in pA626. Since this did not

prevent the gene from promoting clamp cell development in transformation assays, it can be concluded that this region is not essential for this function. This result is not unexpected since it has been shown previously that the A42 b1-1 protein can be truncated by 40 amino acids without loss of function in transformation assays (Tymon, Kües et al. 1992).

```

      ctcgagcatgtatcgattgcatgagaatgtttggagatctttgagggacaatgaatgctg
1  -----+-----+-----+-----+-----+-----+ 60

      tcacctgtggggaagtctttcgaactccgacataaccagtgtctcgcattgttgaagttg
61  -----+-----+-----+-----+-----+-----+ 120

                                     M A I A
      agttcgcgtcgtctcatcgagattcccatccctcctgcatctcaccgcaATGGCTATTGC
121 -----+-----+-----+-----+-----+-----+ 180

      T D V L S E E N T N F I V R S T I S S L
      TACAGATGTTCTCTCGGAAGAGAACACCAATTTTCATTGTGAGGAGCACCATTAGCAGCCT
181 -----+-----+-----+-----+-----+-----+ 240

      R K D F F S L L R G E S T S F D A F L S
      TCGCAAAGATTTCTTCTCCCTCCTCCGTGGGGAATCGACCTCCTTTGATGCTTTCCTCAG
241 -----+-----+-----+-----+-----+-----+ 300

      A Y L K F D S L V L S C Q E S L T D E T
      CGCCTACCTGAAATTCGACTCTCTTGTCTGTCCTGCCAGGAATCGCTCACCGATGAAAC
301 -----+-----+-----+-----+-----+-----+ 360

      L D L V D D F V A P L S T L S S N M V E
      ATTGGACCTAGTGGACGATTTTGTGGCGCCACTATCGACCCTTTCGAGCAATATGGTCGA
361 -----+-----+-----+-----+-----+-----+ 420

      L D A K K E A A S G E F S S N L V G L F
      GCTGGACGCAAAGAAAGAGGCTGCATCCGGCGAATTCTCATCGAATCTCGTTGGCCTCTT
421 -----+-----+-----+-----+-----+-----+ 480

      S T Q L N V V D D Y N
      CTCCACTCAACTGAATGTCGTTGATGACTATAACGgtgggtatttctgtctcaatttaca
481 -----+-----+-----+-----+-----+-----+ 540

                                     D E S T P S Y L E P C A R
      gatgcgctgctgaaaatgttatagATGAATCCACTCCGTCATACCTAGAGCCATGTGCAC
541 -----+-----+-----+-----+-----+-----+ 600

      W L K D N C H N P Y P S A E V R S E I A
      GGTGGCTCAAGGATAATTGTCACAACCCCTACCCAGCGCGGAAGTTCGTTTCAGAAATCG
601 -----+-----+-----+-----+-----+-----+ 660

```

```

      N Q N S N S V K D I D A W F I D A R R R
CAAACCAGAACTCGAACTCCGTGAAGGACATCGATGCTTGGTTTATTGATGCTCGCAGGC
661 -----+-----+-----+-----+-----+-----+ 720

      I G W N D L R R K H F D N K R A K I V Q
GTATCGGCTGGAACGATCTGCGCCGCAAGCATTTCGACAATAAACGAGCGAAGATAGTCC
721 -----+-----+-----+-----+-----+-----+ 780

      A A S I F F N S D L V P L E S D V L Q A
AAGCTGCTTCTATCTTTTTCAACTCAGACCTGGTCCCGTTGGAATCAGATGTCTTGCAGG
781 -----+-----+-----+-----+-----+-----+ 840

      H I E R E F A A I L G R A T T F Y D Q K
CCCATATCGAAAGGGAATTGCGCCGCCATTCTTGGACGTGCCACGACCTTCTACGACCAGA
841 -----+-----+-----+-----+-----+-----+ 900

      F S Q S K L A D K L D S A V K A M T P S
AGTTCTCGCAGAGCAAACCTCGCTGACAAGTTGGATTCTGCTGTCAAGGCAATGACTCCGT
901 -----+-----+-----+-----+-----+-----+ 960

      I R E A L K K E K A S E S Q E D I D S R
CCATCCGGGAGGCTCTGAAGAAAGAGAAGGCTAGCGAAAGTCAGGAAGACATAGATTCTC
961 -----+-----+-----+-----+-----+-----+ 1020

      M S K R A R H A Y P T P E R L P A S P A -
GCATGAGCAAACGCGCCCGTCACGCTTATCCCACTCCTGAACGCTTGCCTGCTTCACCTG
1021 -----+-----+-----+-----+-----+-----+ 1080

                                     |-> start
      E L L T S P P S H V I E L N I D T S P T
CCGAGCTTTTGACCTCCCCACCGTCACATGTCATCGAGTTGAACATCGACACGTCACCGA
1081 -----+-----+-----+-----+-----+-----+ 1140

of cDNA clone
      L S R K R R R P L H S D E D T S L P L G
CCCTGAGTCGCAAGCGTCGTCGACCATTGCACAGTGATGAAGATACCTCGCTCCCGCTTG
1141 -----+-----+-----+-----+-----+-----+ 1200

      K R P R
GAAAACGACCCCGgtatgtaccctctacctgtctgggctttcgctaattcttcgctgcagA
1201 -----+-----+-----+-----+-----+-----+ 1260

      D L S I H R D L S P L K G L P S P A S S
GATCTGTCCATTCATCGTGATCTCAGTCCCCTGAAAGGACTGCCCTCACCAGCCTCCTCC
1261 -----+-----+-----+-----+-----+-----+ 1320

      I Q E E L L E S S P V A S S Q P S S L T
ATTCAAGAGGAGCTTTTAGAATCTTCCCCGTTGCCTCATCTCAACCGTCGTCTCTCACC
1321 -----+-----+-----+-----+-----+-----+ 1380

      P T R S S T C T G K R K R C L S D G F Q
CCGACCCGGTCGTCGACATGCACGGGCAAGCGGAAACGATGCTTATCCGATGGCTTCCAG
1381 -----+-----+-----+-----+-----+-----+ 1440

```

Y P A A K H H R S P Q A V S D P F P A A
TATCCTGCTGCAAAGCACACAGGTCGCCACAGGCCGTTTCGGATCCTTTCCCCGCTGCC
1441 -----+-----+-----+-----+-----+-----+ 1500

S G Q D W D Q W F R E H V L S S P E L T
AGTGGGCAAGACTGGGACCAATGGTTCCGTGAACATGTTCTCTCCTCTCCTGAGCTGACC
1501 -----+-----+C-----+-----+-----+-----+ 1560
| -> C in cDNA -> change of AA from Q->H

L T G D I H P P L T L E A P D S N T P L
CTTACTGGAGACATTCACCCACCGCTTACGCTCGAGGCCCGGACTCCAACACCCCTCTC
1561 -----+-----+-----+-----+-----+-----+ 1620

D I Q L F N F P L I P D L P P S V L P T
GATATTCAACTGTTCAATTTCCCTCTCATCCCTGATCTCCCTCCCTCGGTTCTGCCGACC
1621 -----+-----+-----+-----+-----+-----+ 1680

A P R
GgtaagacatcttgcttgtctattgtgtcaccctgactgactactacggtagCACCACGA
1681 -----+-----+-----+-----+-----+-----+ 1740

A E P N L I E Q L E V P V D K Q V K V D
GCTGAGCCTAATCTTATCGAGCAACTTGAAGTACCGGTCGACAAACAGGTCAAAGTTGAC
1741 -----+-----+-----+-----+-----+-----+ 1800

P D V P T L D Q N L S W M A N D F P P L
CCTGACGTACCAACCCTCGATCAAAACCTCTCGTGGATGGCCAATGATTTTCCTCCACTG
1801 -----+-----+-----+-----+-----+-----+ 1860

L Q P
CTTCgtaagtcctgtgatctgcagtgaaccaaccttaactgataccactcactcagAACCC
1861 -----+-----+-----+-----+-----+-----+ 1920

T N T L P S S A N I Y L H S T P V P T L
GACCAATACACTCCCAAGCTCAGCCAACATTTATTTGCATTCAACTCCTGTCCCAACATT
1921 -----+-----+-----+-----+-----+-----+ 1980

D G I S Q P F A D T Q S S A F L H D P S
GGATGGTATATCGCAACCTTTTCGCGGACACTCAGAGCTCTGCTTTCTCCTCCATGACCCATC
1981 -----+-----+-----+-----+-----+-----+ 2040

L W S N I S D P D L D F S G L F G Q C S
GCTATGGTCCAACATCTCTGATCCTGATCTCGACTTCTCTGGCTTGTTTGGCCAGTGCTC
2041 -----+-----+-----+-----+-----+-----+ 2100

T G T A V A A S I Q V P S H P M L S T S
GACGGGTACTGCCGTGGCCGCGTCGATTCAAGTTCCTTCGCATCCGATGTTGTCGACATC
2101 -----+-----+-----+-----+-----+-----+ 2160

K A L S E Q E R E A K R R E L E E L E A
CAAAGCGTTGTCTGAACAGGAACGTGAAGCCAAGCGCAGGGAGCTCGAGGAACCTTGAAGC
2161 -----+-----+-----+-----+-----+-----+ 2220

```

      R  A  Q  A  I  R  A  E  I  S  A  P  *
      GAGAGCCCAAGCCATTCGCGCCGAAATCTCCGCCCTtgaacctttgctggacctacgct
2221 -----+-----+-----+-----+-----+-----+ 2280

      gctatataccacgctcggtgataccctccatgctttgctgctgctatgctttcgatcttc
2281 -----+-----+-----+-----+-----+-----+ 2340

      gctttcgatcttcattttgtcgcacttgtttcgcaaaccaaactcctcgcaattgtattc
2341 -----+-----+-----+-----+-----+-----+ 2400

      tagccatccagcatactatcgtcattgtgcctttgattatcctattactctagtgtacca
2401 -----+-----+-----+-----+-----+-----+ 2460

                                                    |-> poly A
      ttagaagttagtagtttagattgttacactatgtcaaacatacctatgcatgcaaactat
2461 -----+-----+-----+-----+-----+-----+ 2520

      gatgctttttgtatctcacttcacgcctgctgagagaacctaggcctttgaggctggaga
2521 -----+-----+-----+-----+-----+-----+ 2580

      agcgagactgactggacgaactggctggtgcttctgcttggaattacagggttcggagtac
2581 -----+-----+-----+-----+-----+-----+ 2640

      tggttttcnnnngccgagttcttagacctccctccatatgcagtgcatacaattactag
2641 -----+-----+-----+-----+-----+-----+ 2700

      gcacaggcatccacgaagcacagtgnnttcggtgcctggtgctaaagaacaatcagtaaag
2701 -----+-----+-----+-----+-----+-----+ 2760

      ggtacgtgaggtacccagcaagtcacatattcgacgtcagggacttcgagagatgcgct
2761 -----+-----+-----+-----+-----+-----+ 2820

      ccacgtattgtaaggcaatccgagtgccaagtgcaggagagaagatgacgactttaaggg
2821 -----+-----+-----+-----+-----+-----+ 2880

      ataagtttagagggtttctgggtcacggtattgcttgtggctagtactacagcctttct
2881 -----+-----+-----+-----+-----+-----+ 2940

      ttcccngtcaaacaggagtggcgttggaagggaangtngtagagagtaataactcgta
2941 -----+-----+-----+-----+-----+-----+ 3000

      actcgag
3001 ----- 3007

```

Figure 4.4: DNA sequence of the *C. cinereus b1-3* gene (pA626 and pA627). The predicted protein sequence is shown in one letter code. The start of the pb1-3cDNA clone and the polyadenylation site are indicated.

```

ctcgagtcggtcaatgtctgaagccggttgagctataagttttacttcaagagcagtct
1 -----+-----+-----+-----+-----+-----+ 60

accaggtacgatgctctctctatatcttttgcgtactcgtaattgagatagccattgcaa
61 -----+-----+-----+-----+-----+-----+ 120

ttgcgccgctattgacattcggttgattcagtaacctgaatcgaaactccgcattgtcgtc
121 -----+-----+-----+-----+-----+-----+ 180

caaacacacggtccatagcaatgtccgcggcggattcgctgtgttgcccctttccaaata
181 -----+-----+-----+-----+-----+-----+ 240

      |-> start of full length cDNA
      |               M Q E R P N G D L
cgtatccccccatcctctaacctgagcccgctATGCAGGAACGACCAAACGGTGACCTAT
241 -----+-----+-----+-----+-----+-----+ 300

                                |-> start of 2nd cDNA
W G Q L S Q M S Q S W A R L A R R R L R
GGGGACAATTGTCTCAAATGTCCCAGTCATGGGCACGGTTAGCACGACGTCGCCTGCGGC
301 -----+-----+-----+-----+-----+-----+ 360

P E S G L E T L T Q R L P L P N H L K S
CAGAGTCTGGCCTAGAAACGCTCACCCAACGGCTTCTCTTCCAAATCACCTCAAATCCT
361 -----+-----+-----+-----+-----+-----+ 420

L S H L Y I F P L P V D F M E F I H Q L
TGTCCCACCTCTACATCTTCCCACTCCCCGTCGACTTCATGGAGTTCATTCACCAACTGC
421 -----+-----+-----+-----+-----+-----+ 480

R L T P A A V G K L T V K I D S A I E Q
GCCTTACACCAGCTGCAGTGGGCAAACCTTACCGTCAAGATCGACAGTGCCATCGAGCAGC
481 -----+-----+-----+-----+-----+-----+ 540

Q R Q A R L K T Y Q E M C I R Y V E I S
AAAGACAAGCTCGGCTCAAGACTTATCAAGAGATGTGCATACGATACGTCGAAATAAGTC
541 -----+-----+-----+-----+-----+-----+ 600

Q Q K S G P I D S R A V R N A R N L C E
AGCAGAAATCAGGCCCCATCGATTCCAGGGCTGTCCGCAATGCTCGCAACCTTTGCGAAC
601 -----+-----+-----+-----+-----+-----+ 660

R S F A Q D L A D I K S A F M D C Y N R
GGTCGTTTGCTCAGGATCTCGCAGACATCAAATCCGCATTTATGGATTGTTACAATCGCT
661 -----+-----+-----+-----+-----+-----+ 720

C V Q A H S R S E G K R A P F N A
GTGTACAAGCCCACTCTCGGTCCGAAGGGAAGCGTGCGCCTTTCAACGCagtaagctatc
721 -----+-----+-----+-----+-----+-----+ 780

                                E Y T P L L E K
gagtcggtcagactctgctgatcctgccgcaatagGAATATACCCCCCTTCTTGAGAAGT
781 -----+-----+-----+-----+-----+-----+ 840

```

Y F E Y N A Y P S A R D R E W L A R K S
ACTTCGAATACAATGCCTATCCATCGGCTCGCGACCGCGAATGGCTTGCGCGGAAATCCA
841 -----+-----+-----+-----+-----+-----+ 900

M M S T R Q I E V W
TGATGTCAACGCGTCAAATCGAAGTTTGGgtaagtcttatacaggcttccaggtgttgac
901 -----+-----+-----+-----+-----+-----+ 960

F Q N H R R R A R R D G I
tcgctgaccccgcccatttagTTCCAAAATCACAGACGACGCGCCCGGAGGGACGGAATC
961 -----+-----+-----+-----+-----+-----+ 1020

H F E R L P M D R V P V D I S L E S L E
CACTTTGAAAGACTGCCCATGGACCGAGTACCAGTGGACATATCACTAGAATCCTTGGA
1021 -----+-----+-----+-----+-----+-----+ 1080

D N V P P F T L P R K S E P S E S R W P
GACAACGTCCCGCCCTTCACACTACCACGCAAATCCGAGCCCAGCGAGTCCCGTTGGCCG
1081 -----+-----+-----+-----+-----+-----+ 1140

T P P
ACGCCTCCTgtaagccggtatcaatgctgtcactacatcttcaatctgacccccacacag
1141 -----+-----+-----+-----+-----+-----+ 1200

P E V H P Q V S L S P T T N P T C P P R
CCAGAAGTCCATCCTCAGGTTTCATTATCGCCGACAACAAACCAACATGTCCTCCCCG
1201 -----+-----+-----+-----+-----+-----+ 1260

P A S I L D D I P A A E H A F P T I F P
CCTGCCTCCATTCTTGATGACATTCCTGCCGCGAAGACGCTTTCCCACTATCTTCCCG
1261 -----+-----+-----+-----+-----+-----+ 1320

K G R S D N D R F P F K H G D F N F P A
AAAGGCAGATCCGACAACGACCGTTTTCCCTTCAAGCATGGAGATTTCAATTTCCCTGCT
1321 -----+-----+-----+-----+-----+-----+ 1380

P V W D R R L E N A V A T L P W P C T V
CCCGTATGGGATCGTCGACTTGAAAACGCTGTCGCGACACTCCCTTGGCCTTGCACCGTC
1381 -----+-----+-----+-----+-----+-----+ 1440

H E L C D A L V A T R I R G P P R D D R
CATGAACTGTGTGATGCACTCGTCGCCACACGTATTCGTGGCCCCGCCCCGAGACGATAGA
1441 -----+-----+-----+-----+-----+-----+ 1500

D N Q P T G P W Y S S R V S G P I K A P
GACAACCAACCAACCGGACCTTGGTATTCCTCACGAGTCTCAGGTCCGATAAAAGCGCCC
1501 -----+-----+-----+-----+-----+-----+ 1560

H P A L I R L P K P S L C L R T A S S A
CATCCGGCCTTAATCCGCCTTCCTAAACCCTCTCTTTGCTTACGAACGGCTTCGTCTGCC
1561 -----+-----+-----+-----+-----+-----+ 1620

S I V R Q S P E Q R R D R R T S T S A P
 TCAATCGTCAGGCAGTCACCGGAACAGCGCCGGGACCGCCGAACCTTCAACTTCTGCACCC
 1621 -----+-----+-----+-----+-----+-----+ 1680

R R S N R V Q Y R P R A H P L R L P P S
 AGACGCTCGAACCGCGTCCAGTATCGGCCTCGCGCTCACCCACTCCGTCTCCCTCCCTCC
 1681 -----+-----+-----+-----+-----+-----+ 1740

L A P L A N E P F P R L S S N S F L S S
 CTAGCACCGCTCGCCAACGAGCCATTCCCTCGCCTATCCTCTAATAGTTTCCTGTCTTCC
 1741 -----+-----+-----+-----+-----+-----+ 1800

V D S T D L E T P D G S L S M G P D L Y
 GTGGATTGACTGACCTGGAACACCTGACGGCTCGCTATCAATGGGGCCCGACCTGTAC
 1801 -----+-----+-----+-----+-----+-----+ 1860

L P G N N H R T A D N H F T R V N P H R
 CTTCCCGGCAACAACCAACCGAACCGCGGACAACCACTTCACGCGCGTCAACCCGCACAGA
 1861 -----+-----+-----+-----+-----+-----+ 1920

K L P S Q S Y Q T P T A F D N W N F P S
 AAGCTCCCTTCCCAAAGTTATCAGACCCCTACTGCATTTGATAACTGGAACCTCCCTTCC
 1921 -----+-----+-----+-----+-----+-----+ 1980

Q P A L R L S F P A Y G L T *
 CAGCCCGCCCTGAGGTTAAGTTTCCCGCGTATGGCTTGACAtagataactccgtccgcg
 1981 -----+-----+-----+-----+-----+-----+ 2040

cgcgctacctgcgactcgcaatgtcgatcgtggcagatgaagggaccagtcggccaatca
 2041 -----+-----+-----+-----+-----+-----+ 2100

aaatatatgtggtcctctacaaggacagtatccaggaacggcaccagtcagttcgactg
 2101 -----+-----+-----+-----+-----+-----+ 2160

tcttacctttgttggtcgtgtatcacgctttctccttaatttttctcctaactgacctcg
 2161 -----+-----+-----+-----+-----+-----+ 2220

ctctttacttttgaactttcgtgcatttcgatacccacgtccaccgcattacattgtatt
 2221 -----+-----+-----+-----+-----+-----+ 2280

cctacgccacgctgccagtgtatctttactgagtatacctagtttcaatattaatacgtt
 2281 -----+-----+-----+-----+-----+-----+ 2340

alternative poly A sites
 |-> |->
 cattgcttacttccaaacaacccgttttcccctaggatactttcaacaggatgccgtca
 2341 -----+-----+-----+-----+-----+-----+ 2400

tccgtcccagctctctgtccatggaagtgctactgtacgcctctgcctaataaagtgtg
 2401 -----+-----+-----+-----+-----+-----+ 2460

catgttacagacatccatgtactcgcccaactttgtacgtatagtatatttgctgaaatc
 2461 -----+-----+-----+-----+-----+-----+ 2520

tcaggtgaaatcaccccagggaaaccagacctttcaagtttcagaaagctt
 2521 -----+-----+-----+-----+-----+-----+ 2571

Figure 4.5: DNA sequence of the *C. cinereus b2-3* gene (clone pA625). The predicted protein sequence is given in one letter code. The start of each cDNA clone and polyadenylation sites are indicated.

4.2.1 Start sites

By aligning the *b1-3* and *b1-1* genes, it is seen that the start codon predicted for *b1-3* is 147 nucleotides further downstream from that predicted for *b1-1* (Tymon, Kües et al. 1992). There is a potential ATG start codon in *b1-3* at a corresponding position to that suggested originally for *b1-1* but multiple stop codons are found in all three reading frames between this and the ATG predicted to be the start codon for *b1-3*. It now seems more likely that a second in frame ATG, 138 bp downstream, maybe the actual start codon for *b1-1*. This is suggested by recent work in which the first ATG of *b1-1* was destroyed through site directed mutagenesis without affecting the ability of *b1-1* to work in transformation assays (Kües 1993). The DNA sequence around the second ATG of *b1-1* is very conserved between *b1-1* and *b1-3* (including the ATG codon). Thus it is likely that both genes start at an equivalent position. The ATG at this position fits the consensus for fungal translation initiation sites $CA^C/A^A/CATGGC$ (Ballance 1991). The sequence for both *b1-1* and *b1-3* is CGCAATGGC, with 8 out of 9 bases of the consensus sequence. However, the one change (G instead of A at the -3 position) is at one of the most important positions of the consensus sequence which may result in a poor translation rate of the *b1* messages.

The first in-frame ATG of the pb2-3cDNA2 clone was assumed to be the start codon used by *b2-3*. This ATG is in a position that corresponds to the start predicted for *b2-1* (Kües 1993).

4.2.2 Introns

The cDNA clone isolated for *b1-3* did not extend to the 5' end of the gene. The first intron of *b1-3* had to be predicted by comparison to *b1-1*. All other *b1-3* and all *b2-3* introns were identified by comparing cDNA and genomic DNA sequences. All introns were at equivalent positions in the A42 and A6 genes. In order to confirm the position of the predicted first intron in *b1-3*, 48 intron sequences from sequenced *C. cinereus* genes were analysed to obtain consensus sequences for the splicing donor, acceptor and lariat sites. The sequences chosen for this analysis were the A mating type genes already sequenced and *C. cinereus* tryptophan synthetase, peroxidase and ras like protein (all sequences obtained from the current EMBL database). The results of this analysis are shown in table 4.3.

Table 4.3: *C. cinereus* consensus sequences for intron splice borders and lariat site (a to f refers to the numbering of the bases within a consensus element). 48 introns were analysed. Numbers indicate the occurrence of a base at a particular position.

	DONOR						LARIAT				ACCEPTOR		
	a	b	c	d	e	f	a	b	c	d	a	b	c
A	-	-	28	23	1	5	1	-	10	47	2	48	-
C	-	-	2	10	1	7	41	1	14	1	25	-	-
G	48	-	15	6	45	4	-	-	21	-	-	-	48
T	-	48	3	9	1	32	6	47	3	-	21	-	-
	COPRINUS CONSENSUS SEQUENCES												
	G	T	R	A _N	G	T	C	T	N	A	Y	A	G

The *C. cinereus* intron consensus sequences obtained by this analysis agree well with consensus splice site sequences reported previously for filamentous fungal genes. They are GTNNG and YNG as the respective donor (5') and acceptor (3') sites, with CTNA for the lariat site (Ballance 1991) (Ballance 1986) (Gurr, Unkles et al. 1987). The predicted first intron in the *b1-3* gene shows a 100% match to the *C. cinereus* consensus site and since *b1-1* has an intron at this position, it was considered good evidence for its presence in *b1-3*. All other introns found in *b1-3* and *b2-3* matched the *C. cinereus* consensus splice site consensus sequences well. All introns were about 50 bp long, a typical size for filamentous fungal genes (Gurr, Unkles et al. 1987).

4.2.3 Promotor sequences

There is no obvious TATA box sequence for either of the A6 *b* genes. A potential CAAT element was found 111 bp upstream of the ATG start codon only for the *b1-3* gene but not for *b2-3*. This sequence contains 6 of the 9 bp of the CAAT consensus motif GGC/TCAATCT according to (Montague 1987). The difficulty of locating promotor elements is a common feature of filamentous fungal genes (Montague 1987) (Ballance 1991) (Gurr, Unkles et al. 1987). With respect to the mating type genes, the same observation has been made for the *A α* genes of *S. commune* (Stankis, Specht et al. 1992). In this context, it may be noted that the *A* genes of *C. cinereus* are transcribed at a very low level. Transcription can only be detected in poly(A)⁺ selected RNA and with radioactive probes labelled to a high specific activity (Kües, Richardson et al. 1992). On average 50,000 cDNA clones had to be screened to obtain a single clone containing an *A* gene cDNA suggesting that the *A* transcripts are very rare.

4.2.4 Polyadenylation

Polyadenylation sites as indicated in figures 4.4 and 4.5 were determined by sequencing the cDNA clones of *b1-3* and *b2-3*. Interestingly, the point of polyadenylation does not seem to be restricted to one site in *C. cinereus*. Alternative polyadenylation sites, separated by 6 bp, were found for *b2-3* and the polyA tail of the one cDNA clone of *b1-3* started 99 nucleotides further downstream than that of the cDNA of the corresponding A42 *b1-1* gene (Tymon, Kües et al. 1992) (see appendix for alignment). This suggests that the *b1* genes also have alternative polyadenylation sites particularly since the 3'ends of the *b1* genes are homologous DNA sequences (*b1-1* and *b1-3* are more than 90% identical at the two polyadenylation regions). One explanation for this lack of polyadenylation site specificity could be the absence of clear eukaryotic polyadenylation signal sequences, which is common for fungal genes (Ballance 1991). Neither the A6 nor the A42 *b* genes possess the AAUAAA signal sequence 10 to 30 bases upstream of the polyadenylation site. All polyA tails start after a conserved YA dinucleotide which could function as a cleavage signal during the processing of the initial transcript. The polyA tails were found to have an average length of 30 bp which is typical for filamentous fungal mRNAs (Gurr, Unkles et al. 1987).

4.3 Analysis of the predicted protein sequences of the *b1-3* and *b2-3* genes

4.3.1 The *b1-3* protein

The predicted protein for the *b1-3* gene is 630 amino acids with a molecular weight of 69,705 daltons. It is rich in hydroxylated (18.4%) and acidic amino acids (14.1%). The predicted pI of 4.88 is very acidic and a reflection of the fact that the overall charge of the *b1-3* protein at neutral pH is -26. The high

content of hydroxylated and acidic amino acids is a common feature of eukaryotic transcriptional activators. The negative charge of the protein is not confined to any particular sequence except for a small region near the C-terminus . The predicted properties of the b1-3 protein agree well with these predicted for the b1-1 protein when translated from the second ATG codon. The b1-1 protein is then 632 amino acids long, has a molecular weight of 69,654 daltons, 19% hydroxylated and 13.8% acidic amino acids with a pI of 4.99 at a net charge of -17.

The predicted b1-3 protein has several characteristics of transcription factors such as potential DNA binding, transactivation and dimerization domains. Residues 117 to 174 comprise a homeodomain-like motif. This region is the most highly conserved amongst the *C. cinereus* HD1 proteins as can be seen in figure 4.7 which shows an alignment of the b1-3, b1-1 and d1-1 proteins. The d1-1 protein is included in this figure to allow comparison of allelic HD1 proteins from the same gene-pair (b1-1 and b1-3) with an HD1 protein of a different gene-pair (d1-1). The b1-3 homeodomain clearly falls into the previously defined HD1 group. The alignment in figure 4.6 shows b1-3/b1-1: 72% identity, 81% similarity; b1-3/d1-1: 67% identity, 79% similarity.

B1-3	MAIATDVLSE	ENTNFIVRST	ISSLRKDRFS	LLRGESTSFD	4				
B1-1	MAISAGTTPAC	DSPDDNIRRA	INTLRADLSA	LLRGESVAYS	4				
D1-1	. . MPLSRVAG	DSVDATLNQT	LNCCLPNDLLA	SLRGDATSKS	3				
B1-3	AFLSAYLKF	SLVLS	SCQESL	TDETL	DLVDD	FVAPL	STLSS	8	
B1-1	AFLSACTK	GHAQ	SC	HGML	SDDT	LDLLYS	FSESL	LALSE	8
D1-1	TASAISKVD	ALVQ	SCQQL	GEGTRDLIS	ASLT	AFLPS			7
B1-3	NMVELDAKKE	AASGEFSSNL	VGLFSTQ . . .	LN	VDDYND .	1			
B1-1	NMALLETKKE	AESNKFETA	EV	MAIS	SDKTS	G	LDLS	DDKNEP	1
D1-1	GIHQAEAE	SS	AVLDD	FSNDV	AGLFD G	LALE	TDA	SPK	1
B1-3	. ESTPSYLEP	CARWLKDNCH	NPYP	SAEVR	SE	EIAN	QNSNSV	1	
B1-1	TSPTPAYVEP	CARWLKDNWY	NPYP	SGEVRT	T	QIAR	QTRTSR	1	
D1-1	EPAEPAYIEP	SCRWLKDNWY	NPYP	SPQVR	SE	SIAK	QTGASR	1	
B1-3	KDIDAWFIDA	RRRIGWN	DLR	RKHFD	NKRAK	IVQ	AASIF	FN	:
B1-1	KDIDAWFIDA	RRRIGWNE	VR	RKHFE	NKRVD	IVR	AASIF	FTG	:
D1-1	KDIDAWFIDA	RKRIGWNE	LR	RRC	FDNKRAN	IVE	EATR	FHQ	:

B1-3	S D L V F L E S D V	L Q A H . . . I E R	E F A A I L G R A T	T F Y D Q K F S Q S	232
B1-1	P Q S I P A E V D A	L P D H . . . I E L	E F A G I L S R A R	S L Y E K F S P S	237
D1-1	G K S L E . L P C	I A H H G P T I E N	D L A S I T S Q A S	A W Y D G R L A Q S	232
B1-3	K L A D K L D S A V	K A M T P S I R E A	L K K E K A S E S Q	E D I D . S R M S K	271
B1-1	K L A V K L D T A V	K D M T P S L K E Q	L K N D E A R R K R	E A S T V G I I N Q	277
D1-1	K L A D K L D P V V	K D M T P A L K E K	I K R D A D E R R R	E E R Q . . . K K K	269
B1-3	R A R H A Y P T P E	R L P A S P A E L L	T S P P S H V I E L	N I D T S P T L S R	311
B1-1	R A R H A Y P T P E	R S P A S A A E L L	A S P P S F A I D .	. S D K L P S V G R	315
D1-1	R K R N P Y P S P E	C S P A S S S G A A	L S P V V P V I D L	T G E S Q C S R P N	309
B1-3	. K R R R P L H S D	E D T S L P L G K R	P R D L S I H R D L	S . . . P L K G L P	347
B1-1	. K R R R S L E S D	E T V S S P L C K R	P R S Q S V F C E L	S . . . P V K G L P	351
D1-1	L K R R A S S C S V	E E Y S C R . . K R	P R G D S S H H D Q	S K S N P E T C L P	347
B1-3	S P A S S I Q E E L	L E S S P V A S S Q	P S . . . S L T P T	R S S T C T G K R K	384
B1-1	S P S P S T Q E L	L E T S A A P S P Q	P S L L P K L T P T	D S A R S T G K R K	391
D1-1	S P A T S A Q E D L	S E A K A A S I P T	L Q L D D S T L E P	I P A S G S N K R K	387
B1-3	R C L S D G F Q Y .		 P	A A K H H R . S P Q	403
B1-1	R R L S D G F Q Y .		 P	A A K R P E I R P Q	411
D1-1	R C L S D G F Q Y L	R Q S D T K S V L R	S Y R S L S L V I V	A G M G R L A S P A	427
B1-3	A V S D P F P A A S	G Q D W D Q W F R E	H V L S S P E L T L	T G D I H P P L T L	443
B1-1	V V S D P F P A T S	S E H W E Q W F R E	H V L S S P E L T L	T G D I P P A V T T	451
D1-1	H V R R P H T S R N	Y S G L Q F R L K L	Q I R I P H W T S K	Y S T S L A V P D L	467
B1-3	E A P D S N T P L D	I Q L F N F P L I P	D L P P S V . L P T	A P R A E P N L I E	482
B1-1	D A P D S N T P L D	I Q L F N F P L I P	D L P P S V P V V P	A P T A E L N I I E	491
D1-1	P A P G S S E S E S	 P L L I	M V A N S K I A L G	D S V A E A I P S F	501
B1-3	Q L E V P V D K Q V	K V D P D V P T L D	Q N L S W H A . N D	F P P L L Q P T N T	521
B1-1	P L E V P A V T Q V	N V D P E A T A L D	H T E S W H A . S D	F P P P L Q S T N T	530
D1-1	P L S A H A S V Q W	S P L L D R R S L P	D T L L E T A V N T	Y N P V L Q . . F D	539
B1-3	L P S S A N I Y L H	S T P V P T L D G I	S Q P F A D T Q S S	A F L H D P S L W S	561
B1-1	F P S S S	. . P F S A L D G M	S L P F P D T R S S	A F L P D P S L W S	563
D1-1	L P P S T D I Q P S	S A T I P P M L S S	S S F D P D T . . .	. F A T D P Q E E D	575
B1-3	 N I S D P D	L D F S G L F G Q C	S T G T A V A A S S	Q V P S H P M L S T	597
B1-1	 N I S D P D	L D F S T V F S Q P	S T N S A M T S S S	Q V P L Q P T W L T	599
D1-1	L F S V P I S A A S	F D L P A F T P Q E	D T D P E T		601
B1-3	S K A L S E Q E R E	A K R R E L E E L E	A R A Q A I R A E I	S A P * 630	
B1-1	S R S L S E Q E R E	A K R K E L E E L E	A R A Q A I R A E I	S A P * 632	
D1-1	. . P L T E A E K E	A K K R E L E E L E	A K V Q A I R A Q I	A R S * 632	

Figure 4.7: Alignment of predicted protein sequences of the *C. cinereus* HD1 proteins b1-3, b1-1 and d1-1 (alignment generated using the GCG programs pileup and prettybox)

Two further regions of high homology between the b1-1 and b1-3 proteins correspond to the putative bipartite nuclear localisation signals (Tymon, Kües et al. 1992). Figure 4.7 shows an alignment of these sequences from b1-1 and b1-3 (amino acids 312-332 and 381-400) together with sequences reported or predicted to function as nuclear localisation signals in other proteins.

B1-3.1	K	R	R	R	P	L	H	S	D	E	D	T	S	L	P	L	G	K	R	P	R
B1-1.1	K	R	R	R	S	L	E	S	D	E	T	V	S	S	P	L	L	K	R	P	R
B1-3.2	K	R	K	R	C	L	.	S	D	.	G	F	Q	Y	P	A	A	K	H	H	R
B1-1.2	K	R	K	R	R	L	.	S	D	.	G	F	Q	Y	P	A	A	K	R	P	E
Myc	.	.	.	.	.	.	.	.	.	.	.	.	.	.	P	A	A	K	R	V	K
H2b	.	.	.	.	.	.	.	.	P	A	E	K	.	K	P	A	A	K	K	T	S
Ust	V	A	K	R	.	.	.	H	P	A	R	K	T	K	P	A	A	K	P	K	S
Xen	K	R	K	T	K	E	E	S	P	L	K	D	.	.	K	D	A	K	K	S	K

Figure 4.8: Comparison of putative nuclear localisation signals in b1-3 and b1-1 with similar sequences in mammalian *c-myc* (Myc), *S. cerevisiae* H2B (H2b), *Ustilago maydis* bE2 (Ust) (Kronstad and Leong 1990) and *Xenopus* N1 (Xen) (references according to Tymon et al)

Many transcription factors are thought to have a short half life *in vivo*. The yeast MAT α 2 protein is unstable *in vivo* (Murphy, Shimizu et al. 1993) and has been shown to be an *in vivo* target for ubiquitinylation which then leads to rapid protein degradation (Chen, Johnson et al. 1993). The role of the *C. cinereus* mating type proteins in determining compatibility would suggest that they should also have a relatively short half life. In this context it is interesting to note that the b1-3 protein possesses a potential PEST sequence motif. PEST sequences are rich in proline, glutamic acid, serine and threonine and they are implicated in protein turnover by cytosolic proteinases called calpains (Rogers, Wells et al. 1986) (Hirai, Kawasaki et al. 1991). PEST sequences have been shown to confer a short half life to proteins. Recent studies demonstrate that calpains are present in cells of lower eukaryotes

such as the slime mould *Dictyostelium discoideum* (Vogt 1992) and recognise a number of transcription factors including members of the POU family of homeodomain proteins as substrates for degradation (Watt and Molloy 1993). It is tempting to speculate that the region from amino acids 347 to 380 of the b1-3 protein functions as a PEST sequence. This region has 14% proline, 9% glutamic acid, 32% serine and 12% threonine a total of 67% PEST amino acids. Significantly, this region is highly conserved in the b1-3, b1-1 and d1-1 proteins which indicates that it has a conserved function.

Another conserved region in the HD1 proteins is the C-terminus. A more detailed analysis of this region, including comparisons to other fungal HD1 proteins is given in chapter 8. A more detailed analysis of other possible functional domains in the b1-3 protein is also given in chapter 8 where conclusions will be drawn not only from sequence analysis but also from experimental data.

4.3.2 The b2-3 protein

The predicted b2-3 protein is 534 amino acids long with a predicted molecular weight of 61,100 dalton. The predicted pI of 10.56 is relatively high and reflects the fact that the overall charge of the protein at neutral pH is +19. The protein is rich in hydroxylated (14.4% S+T) and basic (16.7% H+K+R) amino acids. These predicted characteristics are in good agreement with those predicted for the A42 b2-1 protein which is 523 amino acids long, with a molecular weight of 59,302, a pI of 10.66, a net charge of 21 and containing 16.2% S+T and 16% H+K+R.

Figure 4.9 shows an alignment of the b2-3, b2-1 and a2-1 proteins. The a2-1 protein has been included in this alignment to allow the comparison of two

allelic HD2 proteins (b2-1 and b2-3) with an HD2 protein from a different gene-pair (a2-1). b2-3 possesses the *C. cinereus* HD2 type homeodomain (amino acids 158-217). The similarity of the proteins is highest in this region (b2-3/b2-1: 88% identity, 91% similarity; b2-3/a2-1: 71% identity, 83% similarity). The region N-terminal to the homeodomain is more conserved between the b2 proteins (38% identity, 53% similarity) than between b2-3 and a2-1 (19% identity, 44% similarity). The region C-terminal to the homeodomain is also more similar between the two b2 proteins (49% identity, 65% similarity) than between the b2-3 and the a2-1 proteins (35% identity, 51% similarity). A more detailed analysis of these differences in similarity between allelic HD2 proteins and HD2 proteins from different gene-pairs is given in chapter 8.

B2-3	M Q E R P N G D L W	G . Q L S Q M S Q S	. W A R L A R R R L	R P E S G L E T L T	38
B2-1	M I G L E N P E L W	G . Q L R R A T G S	G L V W L	. S V G A K T K L L	33
A2-1	M A G P S R A A K E	A S S A R K I A N T	S R D L M A . . . F	L G A S R T R T L P	37
B2-3	Q R L P L P N H L K	S L S H L Y I F P L	P V D F M E F I H Q	L R L T P A A V G K	78
B2-1	F R P A G H R V L N	S S S S P Q I F P L	L L S F T Q F V S S	L H L T P S A S F Q	73
A2-1	R R S S T A I E I P	P L N L P L	P E E I L D V V L A	Y D V P P A V M F G	73
B2-3	L T V K I D S A I E	Q Q R Q A R L K T Y	Q E M C I R Y V E I	 S Q Q K S	113
B2-1	L V L K V D K V I E	E Q R T L H L T N Y	Q R M C G G Y L E L	 S K Q Q G	108
A2-1	V Q Q T L R N A L L	Q L Q T E C K N K Y	I K T C Q A L S D L	P D T S A S R H R L	113
B2-3	G P S D S R	. A V R N A R N L C	E R S F A Q D L A D	Y K G A F M D C Y N	148
B2-1	M E S D P I	. S V R N M Q T L V	E R S F Q Q L I S R	Y Q G E I M E . . .	140
A2-1	Q Q S N K T F T K L	Y I T R H A P R L R	E Q V I A Q L E T K	L A A P	147
B2-3	R C V Q A H S R S E	G K R A P F N A E Y	T P L L E K Y F E Y	N A Y P S A R D R E	188
B2-1	R Y Q Q T R Q Q L E	V K R T P F N S E Y	T P L L E K Y F E Y	N A Y P S A R D R E	180
A2-1	 P	S K K Q A F N V H Y	I P V L E K Y F E Y	N A Y P T A Q D R A	178
B2-3	W L A R K S M M S T	R Q I E V W F Q N H	R R R A R R L G I H	F E R L P M D R V P	228
B2-1	W L A R K T M M S V	R Q I E V W F Q N H	R R R A R K E G I H	F E R L P M D R V P	220
A2-1	L L A R K S M M S A	R Q I E V W F Q N H	R R R A R K E G K P	L Q R P S P D T L P	218
B2-3	V D I S L E S L E D	N V P P F T L P R K	S E P S E S	. . . R W P T P P P	261
B2-1	V E L S L E S L E K	K L S P L T L P H K	V E L Y G S	. . . R L P T P P P	253
A2-1	V Q I C L S L E Q	E M A P F . V S H K	R E R C P S A D L N	P R T R F P T P P A	257
B2-3	E V H P Q V S L S P	T T N P T C P P R P	A S I L D D I P A A	E H A F P T I F P K	301
B2-1	D A N H I V S P C I	A K V S F D Q P R A	H S I L D R A P G A	R H A P T V Y R K	293
A2-1	E P S P S T S R S G	S K . G P Q H P K Q	D N L D H A I S P	M Y A F P T K F S E	296
B2-3	. G R S D N D R F P	F K H G D F N F P A	P V D R R L E N A	V A T L F W P C T V	340
B2-1	. D G H N D D L F P	C K R G N F N F P A	P V D R C . S S T	T P F K P R H C T M	331
A2-1	L P N I G V D P F P	C K Q G P I S L P A	P V A S R . . S H	T R F Q R E F C S V	334

B2-3	H E L C D A V A T	R I R G P P R D D R	D N Q P T G P V Y S	S R V S G P I K A P	380
B2-1	D Q L C Q T F A A T	R I R G P S K D Q E	D P G Q P G P V Y A	S R V T G P I I A P	371
A2-1	T E L C D A F G M M	R I A G S F . . . S	K A S S S S P V Y S	S R C N G P I V A P	371
B2-3	H P A L I R L P K P	S L C L . R T A S S	A S I V R Q S P E Q	R R D R R T S T S A	419
B2-1	H P A A I R P S E P	R L R L Q K S V G S	P R P S R Q S E S P	R R E R Q V S T P A	411
A2-1	L P A L I R P S L P	R P S L I V E S S E	A S R S K S P . . .	. . . R S V S S S C	405
B2-3	P R R S N R V Q Y R	P R A H P L R L P P	S L A . P L A N E P	F P R L S S H S F L	458
B2-1	P K P S N R R Q . E	P R S H T T T P S P	S L P T T F H K R T	I P R L S S Q S S L	450
A2-1	V R T K Q C R Q S T	L R T C L T G C F A	 Q F G K K S	L A R A S S N S S L	441
B2-3	S S V D S . T D L E	T P D G S L S M G P	D L Y L P G N N H R	T A D N H F T R V N	497
B2-1	S F V D S . S D L D	T P D G S P S M S P	A S A L P P I V H E	H I G D P Y A V I N	489
A2-1	S S Q G S W S D L E	T P S S P N P S S	T I G L P S V Y G V	D E V H F A P M S D	481
B2-3	P H R K L P S Q S Y	Q T P T A F D N W N	F P S Q P A L R L S	. . F P A Y G L T *	534
B2-1	. Q D K A T S P A I	Y L Q T P F N Y W R	L P S R P P L G L N	S L F M A . . . *	523
A2-1	P S H Q A S S V A F	E S Y A A C P K T	F T E A L S W S L S	T R H S A S L L P *	520

Figure 4.9: Alignment of predicted protein sequences of the *C. cinereus* HD2 proteins b2-3, b2-1 and a2-1 (alignment generated using the GCG programs pileup and prettybox)

The b2-3 protein like b1-3, also has a possible PEST sequence. The region from amino acids 450 to 477 contains 3 prolines, 1 glutamic acid, 8 serine and 2 threonines, a total of 52% PEST residues. This sequence is the most highly conserved part of the C-terminal region of the proteins (75% identity, 82% similarity between b2-3 and b2-1) and it is tempting to speculate that it has a conserved function.

No bipartite nuclear localisation signal similar to those predicted for b1-1/b1-3 can be found in the b2-1/b2-3 proteins. Clusters of positively charged residues could, however, function as nuclear targeting signals (Goldfarb, Goneli et al. 1992) (Tapscott, Davis et al. 1988). A compatible interaction of the A proteins is assumed to involve dimerization of compatible HD1 and HD2 proteins (Kües and Casselton 1992). One can speculate that the HD2 proteins can only

enter the nucleus once they have successfully dimerised with a compatible HD1 protein. Thus HD2 proteins might be restricted to the cytosol of the monokaryotic cells and this may be important for the initial recognition of a compatible HD1 protein following hyphal fusion.

4.4 Detailed analysis of the homeodomains

Database searches with the b1-3 protein sequence yield a limited number of alignments with homeodomain proteins. This reflects the fact that the b1-3 HD1 motif is a non-typical homeodomain. The most striking feature is the replacement of the "invariable" N in the recognition helix (helix 3) by a D (amino acid 164 in b1-3 sequence in figure 4.7) in the *C. cinereus* domain. Although this exchange can be considered to be a conservative replacement (Taylor 1986) it is difficult to imagine how D can provide the DNA binding function that has been attributed to the N in the homeodomain consensus sequence (position 51 in homeodomain consensus sequence, numbering according to the conventions used in (Quian, Billeter et al. 1989) and (Kissinger, Liu et al. 1990)) (Wolberger, Vershon et al. 1991) (Pabo and Sauer 1992).

X-ray diffraction studies of co-crystallised homeodomain/DNA binding site complexes show that the N (= asparagine) in the consensus sequence binds to an adenine base in the DNA-binding site via two hydrogen bonds donating one H of the amido group and accepting one H from the N₆ position of the adenine ring (Pabo and Sauer 1992). The carboxyl group of the D (= aspartic acid) would normally be in its salt state (deprotonated) in aqueous solution and could not form these hydrogen bonds. It is possible that the HD1 homeodomain does not provide the essential DNA contact mediated in a typical homeodomain by the asparagine. HD1 proteins contain a high proportion of acidic side chains and the microenvironment created by the

three dimensional structure of the protein might change the pK_a of the carboxyl side chain of the aspartic acid residue sufficiently to allow protonation of this group. This might then allow formation of two hydrogen bonds similar to the situation in MAT α 2, one of the homeodomain proteins that has been co-crystallised with its DNA target site (Wolberger, Vershon et al. 1991). Several reports describing the three dimensional structure of DNA binding proteins bound to their DNA recognition sequences show that water molecules may form a bridge between an amino acid and a DNA base by providing a hydrogen bond (Pabo and Sauer 1992). This might be an alternative mode of function for the D in the HD1 recognition helix. Figure 4.10 shows an alignment of the HD1 homeodomain from the *C. cinereus* b1-3, b1-1 and d1-1 proteins to some fungal homeodomains of the HD1 type (Kües and Casselton 1992) and to some homeodomain sequences that were found using the BLAST homology search program.

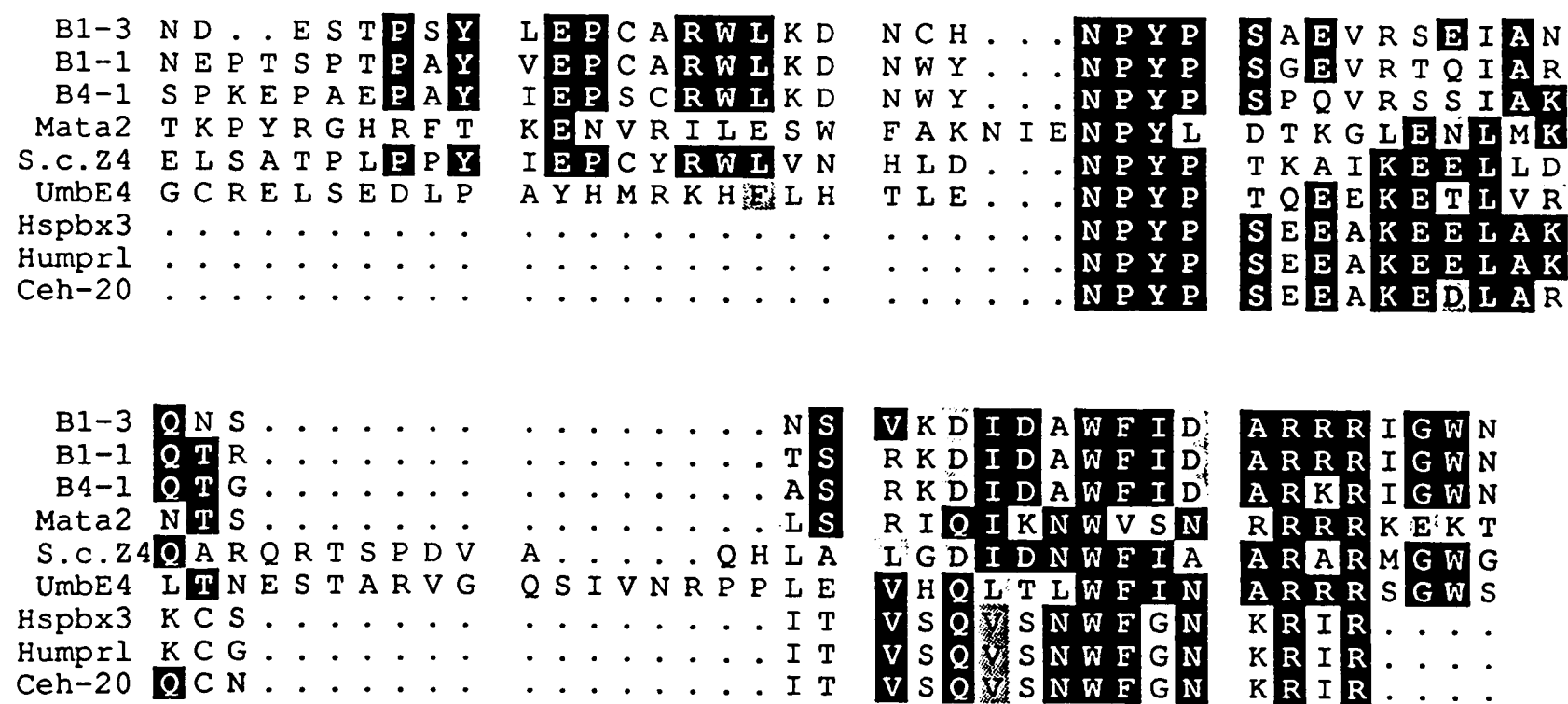


Figure 4.10: Alignment of the b1-3 homeodomain sequence to some fungal homeodomains (B1-1=b1-1 and B4-1=d1-1 from *C. cinereus*, Mata2=MAT α 2 from *S. cerevisiae*, S.C.Z4=*S. commune* Z4, UmbE4=*U. maydis* bE4)and some homeodomain sequences obtained by database search using the BLAST algorithm(accession numbers: Hspbx3=pir S19010, Humprl=pir B33061, Ceh-20=gp U01303)

Figure 4.11 shows an alignment of the HD2 homeodomain of b2-3, b2-1 and a2-1 to other fungal homeodomains of the HD2 type and to some homeodomains from higher eukaryotes obtained by screening the databases. It reveals that there is similarity to classical homeodomains throughout the whole length of the HD2 domain. The HD2 domain also contains all residues considered to be invariable for typical homeodomains (WF_N_R= amino acids 48 to 53 in the homeodomain consensus sequence). Database searches with the b2-3 protein identify more than 450 of the 500 alignments reported to homeodomain proteins. The *C. cinereus* HD2 homeodomain is thus a typical homeodomain as found in proteins isolated from a wide spectrum of eukaryotic organisms.

3-dimensional structures have been obtained for the homeodomains of the engrailed (Kissinger, Liu et al. 1990), the *S. cerevisiae* MAT α 2 (Wolberger, Vershon et al. 1991) (Phillips, Vershon et al. 1991), the HNF1/LFB1 (Ceska, Lamers et al. 1993) and the Oct3 (Morita, Shirakawa et al. 1993) proteins. Engrailed and MAT α 2 were cocrystallised with their DNA binding sites and these studies identified a few key residues involved in making contacts with the DNA target site. Comparisons to the b1-3 and b2-3 homeodomains should therefore help to establish whether the b1-3 HD1 homeodomain shows any more unusual features. Table 4.4 summarises the results of this comparison. Numbering is according to the conventions used by (Quian, Billeter et al. 1989) and (Kissinger, Liu et al. 1990).

2-3prot	S E G K R A P E N A	E Y T P L L E K Y F	E Y N A Y P S A R D	R E W L A R K S M M
2-1prot	L E V K R T P E N S	E Y T P L L E K Y F	E Y N A Y P S A R D	R E W L A R K T M M
2-1prot	P P S K K Q A E N V	H Y I P V L E K Y F	E Y N A Y P I A Q D	R A L L A R K S M M
Mata1	S P K G K S S I S P	Q A R A F L E Q V F	R R K Q S L N S K E	K E E V A K K C G I
Y1prot	S K K P R P K E H S	E Y T P L L E L Y F	R F N A Y P T Y A D	R R V L A E K T G M
bw2prot	P L K T G R G H D S	E A V R I L E Q A F	K H S P N I T P A E	K F R L S E V T G L
xc4_Rat		 L E K E F	H Y N R Y L T R R R	R I E I A H S L C L
m07_Cae	. P R R R T T E T V	E Q L Y L L E M Y F	A Q S Q Y V G C D E	R E R L A R I L S L
xb9_Mou	S R K K R C P Y T K	Y Q T L E L E K E F	L F N M Y L T R D R	R H E V A R L L N L
xb5_Chi	. K R A R T A Y T R	Y Q T L E L E K E F	H F N R Y L T R R R	R I E I A H A L C L
xb9_Hum	S R K K R C P Y T K	Y Q T L E L E K E F	L F N M Y L T R D R	R H E V A R L L N L
m90_Api	G . . . R Q T Y T R	Y Q T L E L E K E F	H Y N R Y L T R R R	R I E I A H A L C L
msc_Api	V K R Q R T S Y T R	Y Q T L E L E K E F	H F N R Y L T R R R	R I E I A H A L C L
xb4_Hum	. K R S R T A Y T R	Q Q V L E L E K E F	H Y N R Y L T R R R	R V E I A H A L C L
men_Sch	E K R P R T A F S G	E Q L A R L K H E F	T E N R Y L T E R R	R Q E L A R E L G L
xa5_Sal	. . G R G Q T Y T R	Y Q T L E L E K E F	H F N R Y L T R R R	R I E M A H A L C L
xc5_Not	. K R S R T S Y T R	Y Q T L E L E K E F	H F N R Y L T R R R	R I E I A N N L C L
xd4_Bra	P K R S R T A Y T R	Q Q V L E L E K E F	H Y N R Y L T R R R	R V E I A H T L C L

2-3prot	S T R Q I E V W F Q	N H R R R A R R D G
2-1prot	S V R Q I E V W F Q	N H R R R A R K E G
2-1prot	S A R Q I E V W F Q	N H R A R A R K E G
Mata1	T P L Q V R V W F I	N K R M R S K . .
Y1prot	L T R Q I T V W F Q	N H R R R A K G P L
bw2prot	K P K Q V T I W F Q	N R R N R K G K K N
Hxc4_Rat	S E R Q I K I W F Q	N R R M K W K K D .
m07_Cae	D E Y Q V K I W F Q	N R R I R M R R E A
xb9_Mou	S E R Q V K I W F Q	N R R M K M K K . .
xb5_Chi	S E R Q I K I W F Q	N R R M K W K K D N
xb9_Hum	S E R Q V K I W F Q	N R R M K M K K . .
m90_Api	T E R Q I K I W F Q	N R R M K W K K E N
msc_Api	T E R Q I K I W F Q	N R R M K W K K E H
xb4_Hum	S E R Q I K I W F Q	N R R M K W K K D .
men_Sch	N E A Q I K I W F Q	N K R A K I K K A S
xa5_Sal	S E R Q I K I W F Q	N R R M K W K K D N
xc5_Not	N E R Q I K I W F Q	N R R M K W K K D S
xd4_Bra	S E R Q I K I W F Q	N R R M K W K K D .

Figure 4.11: alignment of the b2-3 homeodomain to other fungal HD2 type homeodomains (B2-3prot, b2-1prot, a2-1prot = *C. cinereus* A factor HD2 proteins; Mata1 = *S. cerevisiae* MATa1 protein; Y1prot = *S. commune* Y1 protein; bw2prot = *U. maydis* bW2 protein) and to some other homeodomain sequences obtained from a database search (alignment generated using the PredictProtein mailserver at EMBL.Heidelberg.DE and the GCG program prettybox, Swissprot database accession numbers: Hxc4_Rat = P18865, Hm07_Cae = P20270, Hxb9_Mou = P20615, Hxb5_Chi = P14838, Hxb9_Hum = P17482, Hm90_Api = P15860, Hmsc_Api = P15859, Hxb4_Hum = P17483, Hmen_Sch = P14150, Hxa5_Sal = P09637, Hxc5_Notvi = P31262, Hxd4_Bra = P22574)

Tables 4.4 shows that the *C. cinereus* HD1 class of homeodomain proteins diverges considerably from other homeodomains at positions which have been proven to mediate the DNA binding function of these proteins and which are generally highly conserved among homeodomain proteins. This divergence is especially noticeable for the residues that bind to the bases of the DNA target sequence. This can be interpreted in several ways: 1) The HD1 homeodomain does not function as a DNA binding domain. 2) The HD1 homeodomain binds DNA relatively non-specifically via sugar phosphate backbone contacts but lacks residues for sequence specific binding to the bases of DNA target sites. 3) The DNA sequence of the target site might differ considerably from the TAAT core motif recognised by typical homeodomains (Hayashi and Scott 1990). The HD1 homeodomain uses different amino acid residues than typical homeodomains to make contacts to the bases of the DNA target site which would suggest a different mechanism of binding. The *C. cinereus* HD2 homeodomain is a typical homeodomain and shares most of the residues implicated in DNA binding by MAT α 2 and *engrailed*.

The next question addressed was whether a prediction of the 3 dimensional structures of the b1-3 and b2-3 homeodomains matches the overall structure of the four homeodomains that have been determined experimentally (Mat α 2 (Phillips, Vershon et al. 1991), *engrailed* (Kissinger, Liu et al. 1990), LFB1/HNF1 (Ceska, Lamers et al. 1993), oct3 (Ceska, Lamers et al. 1993)). All four homeodomains show a non-structured N-terminus, followed by a helix-turn-helix-loop-helix motif.

Table 4.4: Conservation of amino acid residues shown to make contacts with the DNA target site for the *engrailed* (Kissinger, Liu et al. 1990) and the MAT α 2 homeodomain (Wolberger, Vershon et al. 1991) (+/- indicates whether an amino acid is conserved, (+) means that a similar amino acid is present. * indicates that the HD1 homeodomain has a conserved F residue which is however at position 7 rather than position 8)

	<i>engrailed</i> Protein/DNA contacts made to												
	a) DNA bases					b) sugar phosphate backbone							
	R3	R5	I47	Q50	N51	T6	F8	Y25	R31	W48	R53	K55	K57
HD1	-	-	-	-	-	-	+	*	+	+	+	(+)	-
HD2	-	+	(+)	+	+	-	+	+	+	+	+	(+)	(+)
oct3	+	+	(+)	-	+	+	-	-	-	+	+	+	+
MAT α 2	-	-	-	-	+	-	+	+	-	+	+	+	+
MATa1	(+)	(+)	(+)	-	+	(+)	(+)	-	-	+	+	(+)	+

	MATα2 Protein/DNA contacts made to											
	a) DNA bases				b) sugar phosphate backbone							
	R7	S50	N51	R54	F8	Y25	R42	Q44	W48	R53	K55	K57
HD1	-	-	-	(+)	+*	+	-	-	+	+	(+)	-
HD2	-	-	+	+	+	+	-	+	+	+	(+)	(+)
oct3	-	-	+	-	-	-	(+)	-	+	+	+	+
en	-	-	+	-	+	+	-	+	+	+	+	+
MATa1	-	-	+	-	-	-	-	+	+	+	(+)	+

The results presented in table 4.5 show that the secondary structure predicted for the b1-3 and the b2-3 proteins agrees well with structures obtained from physicochemical analysis of the other homeodomain proteins listed in the table. Despite the atypical amino acid sequence of the HD1 homeodomain, it appears from this analysis that it can take up the secondary structure of a typical homeodomain and may bind DNA in a similar way. Nevertheless, the variation in the amino acid residues predicts a DNA binding sequence that diverges from that of the classical homeodomain TAAT core motif (see above).

Table 4.5: Comparison of the helix positions predicted for b1-3 and b2-3 (using the PredictProtein electronic mail server at EMBL-Heidelberg.DE, (Rost, Schneider et al. 1993)) with helices determined experimentally for MAT α 2 (Wolberger, Vershon et al. 1991), *engrailed* (Kissinger, Liu et al. 1990), oct3 (Morita, Shirakawa et al. 1993) and LFB1/HNF1 (Ceska, Lamers et al. 1993) (LFB1 has a 21 amino acids insertion between helices 2 and 3 at position 38, numbered 38_1 to 38_21). Numbering according to the conventions used in (Quian, Billeter et al. 1989) and (Kissinger, Liu et al. 1990).

	helix 1	helix 2	helix 3
MAT α 2	10 - 22	28 - 38	42 - 58
<i>engrailed</i>	10 - 22	28 - 38	42 - 58
oct 3	10 - 21	28 - 37	42 - 53
LFB1/HNF1	10 - 21	28 - 38_8	42 - 57
b1-3 (predicted)	11 - 18	28 - 35	42 - 56
b2-3 (predicted)	10 - 22	28 - 37	43 - 57

4.5 Construction of chimeric *b* genes of the *C. cinereus* A42 and A6 mating type factors

An important feature of the A proteins is the ability to distinguish compatible partners. b1-1 and b2-3 can interact to promote sexual development whereas b1-1 and b2-1 cannot. For the similar proteins specified by the *b* mating type genes of *U. maydis*, it has been shown that the allele specificity is encoded in the 5' part of the genes. The homeodomain and the region C-terminal to it is homologous sequence in all alleles (Gillissen, Bergemann et al. 1992) (Kronstad and Leong 1990).

Compared to the *U. maydis* genes, the *A* genes from *C. cinereus* have little DNA sequence conservation, but nevertheless the region 5' to the homeodomain is the least conserved between alleles (see data in section 4.4). With the complete sequences of *b1-1*, *b1-3* and *b2-1*, *b2-3* available, it was decided to test whether the 5' part of the *C. cinereus* *A* genes conferred allele specificity. Chimeric genes were constructed *in vitro* by fusing the 5' half of the *b1-3* gene to the 3' half of the *b1-1* gene and by fusing the 5' half of the *b2-1* gene to the 3' half of the *b2-3* gene. The chimeric *b1* gene was designated *b1-3/1* and prepared using a common *Cla*I site within the homeobox sequences of *b1-1* and *b1-3*. The cloning strategy is illustrated in figure 4.12. The hybrid gene was derived from the clone pAMT3, which contains a *b1-1* gene lacking 40 codons from the 3' terminus. This truncated gene has been shown to function in transformation assays (Tymon, Kües et al. 1992), and it was therefore assumed that this truncation would not affect the function of the hybrid gene.

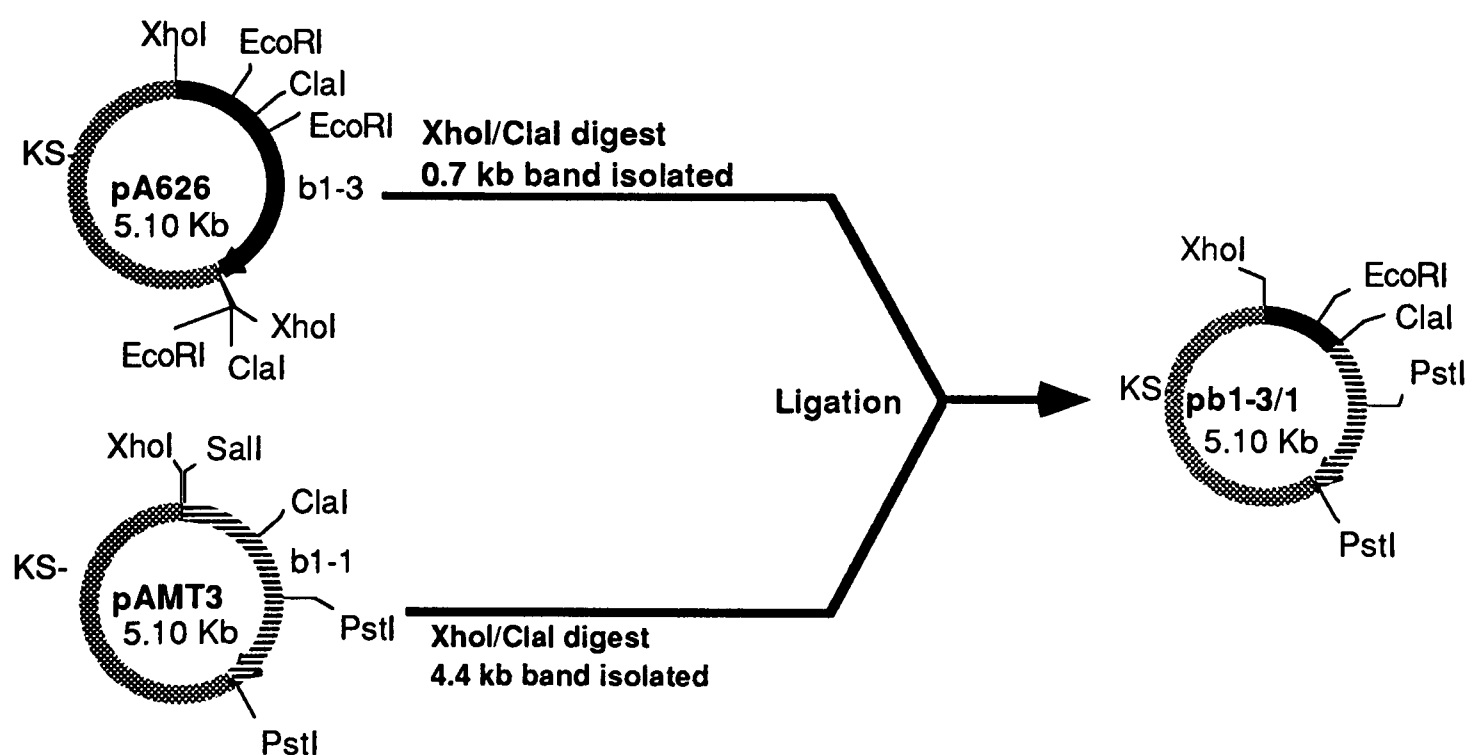


Figure 4.12: Cloning strategy for the preparation of the *b1* chimeric gene construct *pb1-3/1*

Similarly, the 5' part of the *b2-1* gene was fused to the 3' part of the *b2-3* gene using a common *Nco*I site immediately 3' to the homeodomain (position 1040 in figure 4.4, 6 codons 3' to the homeodomain sequence). The cloning strategy is illustrated in figure 4.13. The hybrid gene was designated *b2-1/3*. pA6225 ends at position 1885 of the sequence shown in figure 4.4 thereby lacking the last 45 codons of the predicted *b2-3* coding sequence. This should not prevent the hybrid gene from being active since it has been shown (Asante-Owusu 1993), that at least 700 nucleotides of HD2 genes can be removed from the 3' end without loss of activity in transformation assays

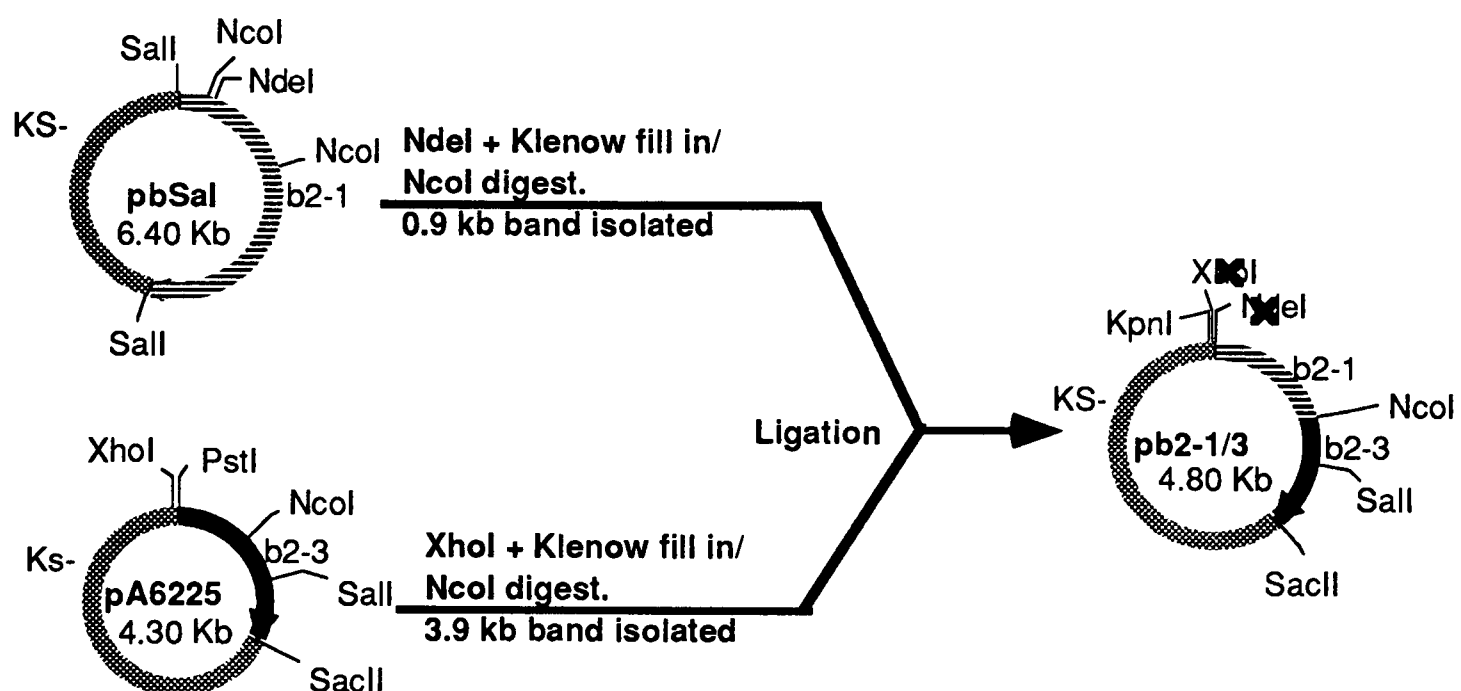


Figure 4.13: Cloning strategy for the construction of the *b2* chimeric construct *pb2-1/3*

- A: λ HindIII size marker
 B: pA626, ClaI/XhoI
 C: pAMT3, ClaI/XhoI
 D: pb1-3/1, ClaI/XhoI
 E: pb1-3/1, EcoRI
 F: λ HindIII size marker
 G: pA6225, NcoI/XhoI
 H: pbSal, NcoI/NdeI
 I: pb2-1/3, KpnI/NcoI
 J: pb2-1/3, KpnI

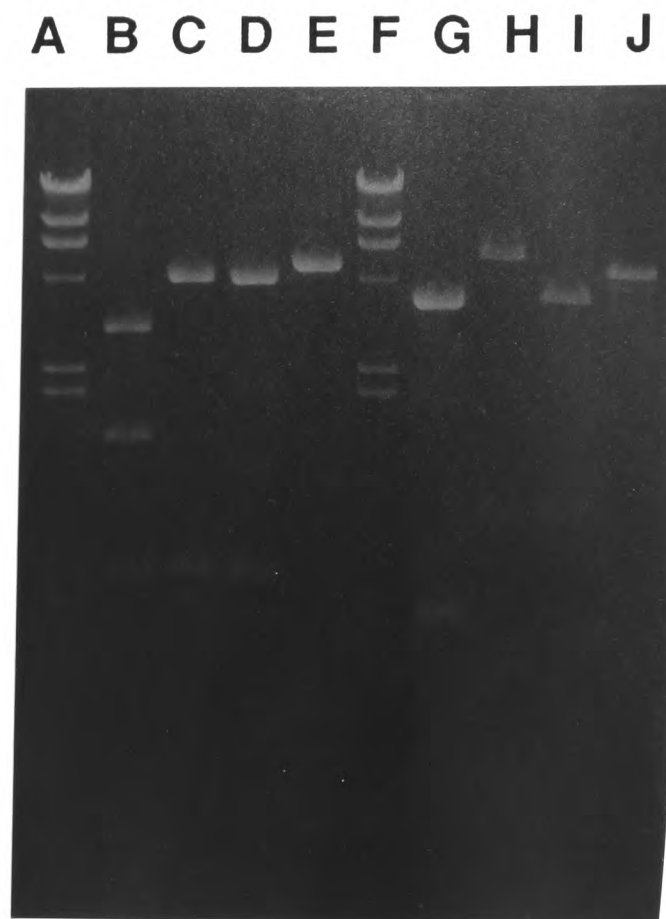


Figure 4.14: Ethidium stained agarose gel showing restriction digests illustrating the cloning strategies used to obtain the chimeric genes *b1-3/1* and *b2-1/3*.

pb1-3/1 and pb2-1/3 together with pA626 and pA625, containing the *b1-3* and the *b2-3* gene respectively, were used to transform an A6 (LT2) and an A42 (LN118) strain of *C. cinereus*. The results of these transformations together with data obtained for the A42 *b1-1* and *b2-1* gene (Kües, Richardson et al. 1992) are summarised in table 4.6. The data show that the allele specificity of the chimeric genes follows the derivation of the 5' portion of the fusion genes. As for the *U maydis* *bE* and *bW* genes, allele specificity of *C. cinereus* A genes is determined by the 5' part of the genes.

Table 4.6: Results of transformation experiments with chimeric and wild type *b* genes of the A42 and A6 mating type factor (+/- indicate formation or absence of clamp cells respectively; where no clamps were detected 100 transformants were screened)

	transformed	into
	LT2 (A6)	LN118 (A42)
<i>b1-1</i>	+	-
<i>b1-3</i>	-	+
<i>b1-3/1</i>	-	+
<i>b2-1</i>	+	-
<i>b2-3</i>	-	+
<i>b2-1/3</i>	+	-

Although the fusion point in the two hybrid genes lies within or immediately behind the homeodomain respectively, it is still apparent that it is the region 5' to the homeodomain that determines allele specificity. The homeodomain sequence and the sequence immediately following it in all *b2* genes is highly conserved (around 90% similarity on amino acid level) and is assumed to bind to an identical DNA target site. This region is, therefore, unlikely to carry the information conferring allele specificity . On the other hand, the region 5' to the homeodomain is the most variable region between genes and alleles and this experiment demonstrates its involvement in determining allele specific interactions between *C. cinereus* *A* genes.

4.6 Mapping the region between the *b1-3* and the *d1-1* genes of the A6 mating type factor

In order to complete the analysis of the A6 β locus, a detailed restriction map was prepared for the region between the *b2-3* and *d1-1* genes. The β locus of the A42 factor contains a divergently transcribed *b* gene pair and two odd HD1 genes, *c1-1* and *d1-1*. The *d1-1* gene is shared between the two A factors but the size of the A6 β locus suggests that there is no room for another specificity gene between *b2-3* and *d1-1*.

This analysis was made using the phage λ 4.1 (figure 3.3). Phage DNA was digested with *Sal*I, *Cla*I, *Cla*I/*Hind*III and *Hind*III and subjected to Southern blot analysis. The gel and corresponding autoradiographs are shown in figure 4.15. The hybridisation probes used were a 0.4 kbp *Hinc*II fragment from the A6 *b2-3* gene (A6219), a 0.8 kbp *Bam*HI/*Sal*I fragment from the 5' end of the A42 *d1-1* gene (0.8kb B/S) and a 1.8 kbp *Hind*III/*Bam*HI fragment from the 3' half of the A6 *d1-1* gene (1.8kb H/B) (the position of these hybridisation probes with respect to a restriction map of the A6 β locus is shown in figure 4.16).

The important observation is that probes derived from the 5' end of *d1-1* (0.8 kb B/S) and from the 3' end of *b2-3* (A6219) hybridised to the same 1.5 kbp *Sal*I fragment. Approximately 1.15 kbp of this fragment derived from the sequenced region of the *b2-3* gene which means that the *d1-1* gene starts within the remaining 0.4 kbp. These additional data allowed the construction of a detailed restriction map of the region between the *b2-3* and the *d1-1* genes as shown in figure 4.16.

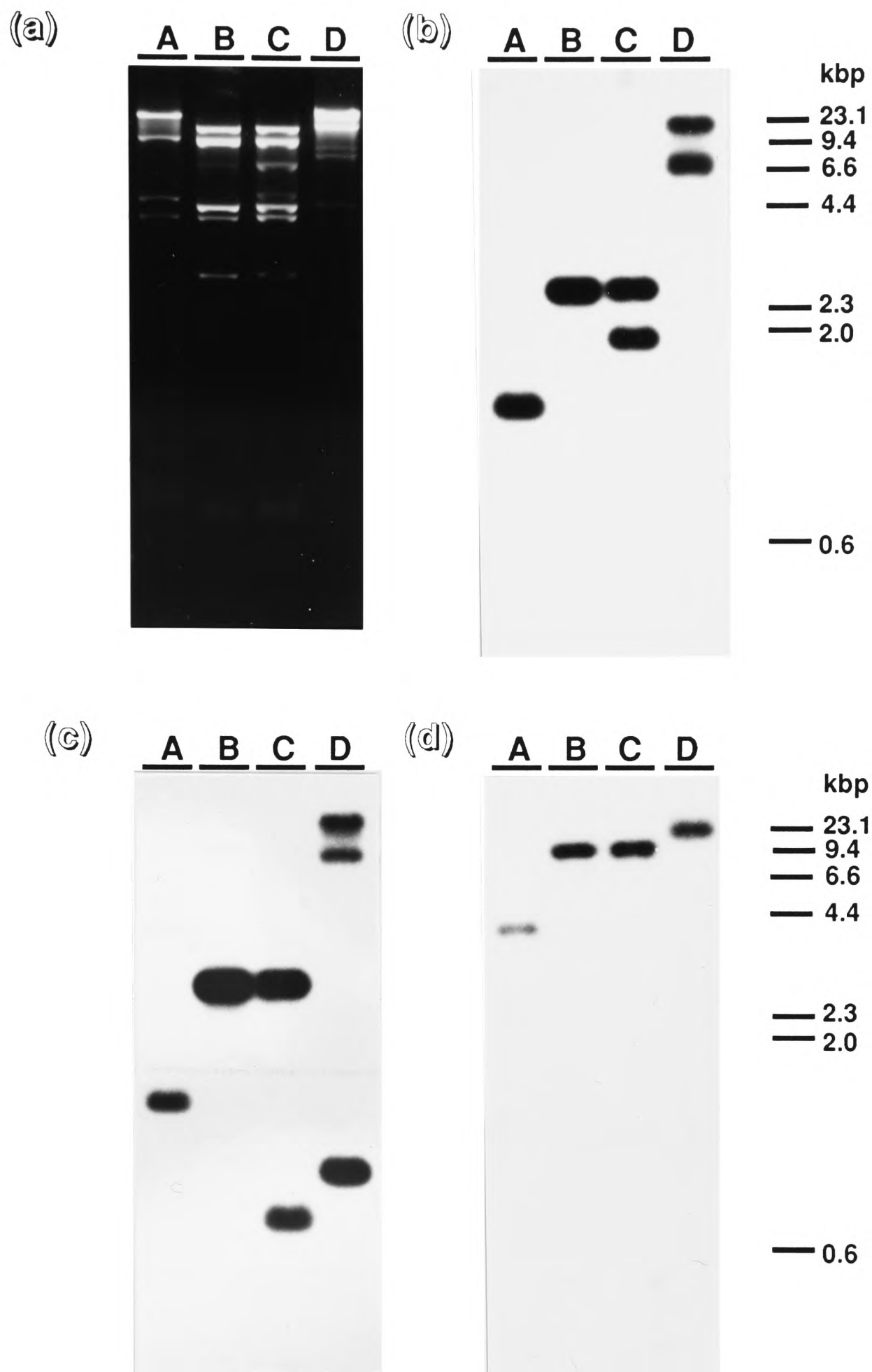


Figure 4.15: Southern blot analysis of phage λ 4.1. (a) ethidium stained agarose gel and the corresponding autoradiographs of three Southern hybridisation experiments (probes A6219 (b), 0.8kbB/S (c) and 1.8kbH/B (d)). A=SalI, B=ClaI, C=ClaI/HindIII, D=HindIII.

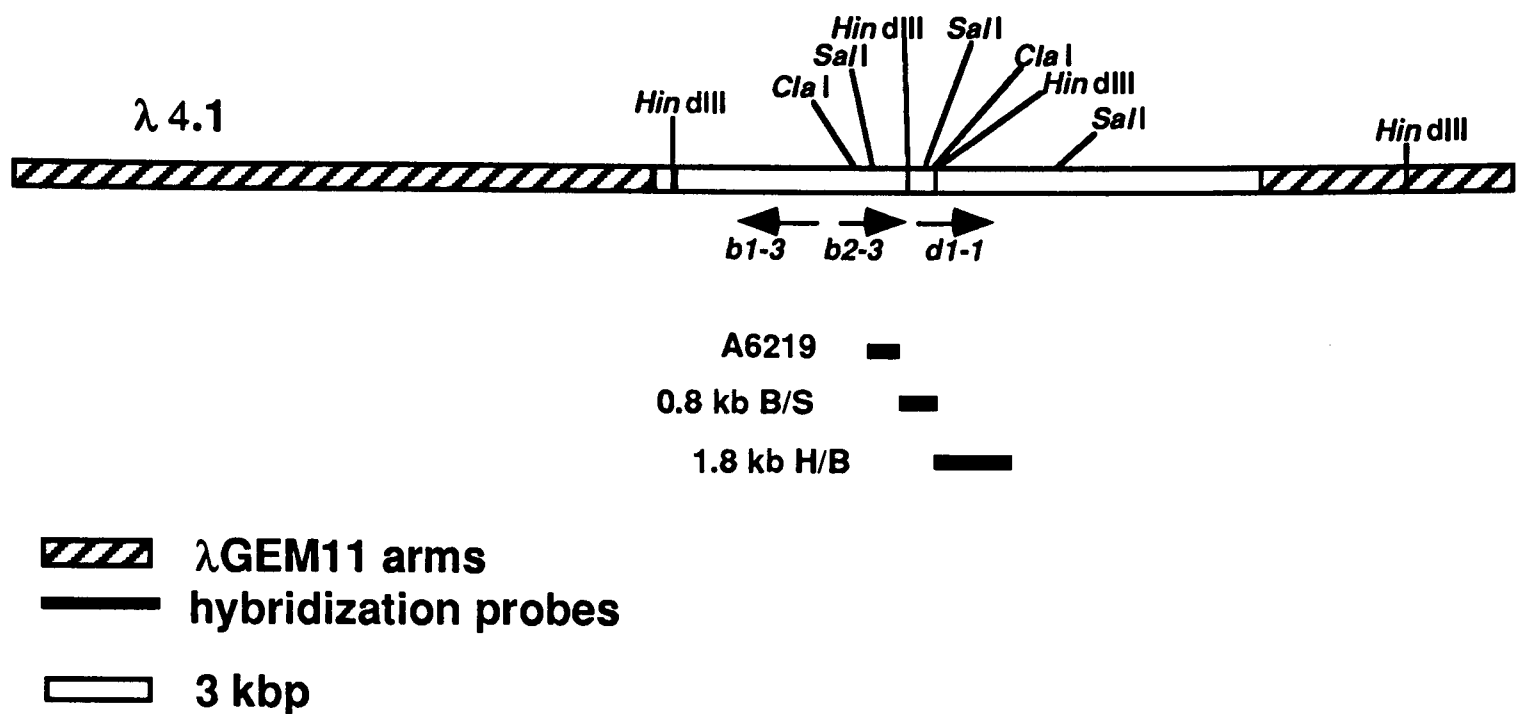
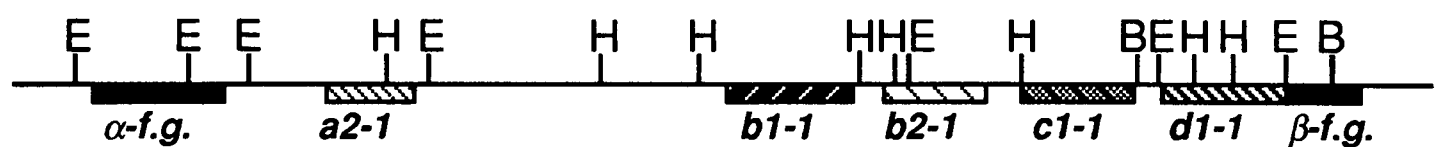


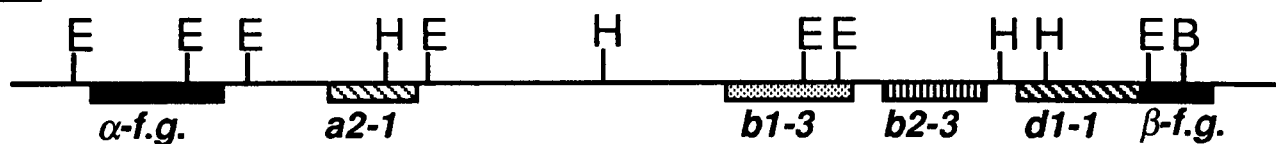
Figure 4.16: Restriction map of phage λ 4.1 showing the position of the genes in the β locus of the A6 mating type factor.

A complete map of the A6 mating type factor can now be presented (figure 4.17). There are only three specificity genes in the A6 β locus, *b1-3*, *b2-3* and *d1-1*. A42 in comparison contains *b1-1*, *b2-1*, *c1-1* and *d1-1* which shows that A mating type factors may have variable numbers of specificity genes.

A42



A6



— 1 kbp

Figure 4.17: Comparison of the A42 and A6 mating type factors of *C. cinereus* showing the position of specificity genes (E=*EcoRI*, H=*HindIII*, B=*BamHI*)

Chapter 5

Raising polyclonal antibodies against the *C. cinereus* b1-1 mating type protein

Antibodies against the A factor proteins of *C. cinereus* would be useful tools with which to characterise the A proteins *in vivo* and *in vitro*. Western blot analysis could be used to monitor the amounts of A proteins present during various stages of sexual development, and also to identify certain types of post-translational modification. Immunofluorescence studies might make it possible to determine the localisation of the A proteins in the cells during various stages of sexual development. Co-immunoprecipitation experiments could identify proteins that interact with the A proteins or interactions between different A proteins. This thesis focuses on the interactions between the proteins of the *b* gene pairs of the A42 and A6 mating type factors. The complete DNA sequences and cloned cDNA were available for both *b1-1* and *b2-1* at the start of this study. This made it possible to prepare plasmid constructs for overexpression of these genes in *E. coli* in order to produce the required amount of protein for immunisation.

5.1 Plasmid constructs for overexpression of the A42 *b* gene-pair in *E. coli* using T7 RNA polymerase systems

5.1.1 Constructs for expression of *b1-1* using the pT7-7 vector system

A gene sequence cloned in the plasmid vector pT7-7 allows high level expression of its protein in *E. coli*. pT7-7 was developed by Tabor and Richardson (Tabor and Richardson 1985) and contains a T7 RNA

polymerase promotor upstream of a multiple cloning site. The T7 RNA promotor is not recognised by the endogenous *E. coli* RNA polymerases and therefore transcription of any cloned genes solely depends on the presence of T7 RNA polymerase. Various experimental systems have been developed to provide an inducible T7 RNA polymerase so that the time of expression of the foreign gene can be regulated. Two systems were used in this study.

The first was a two plasmid system in which one plasmid contained the gene for the T7 RNA polymerase and the second plasmid, a pT7-7 construct, contained the gene to be expressed. The plasmid with the T7 RNA polymerase gene (pGP1-7) also contained the kanamycin resistance gene whereas pT7-7 carries the ampicillin resistance gene. It is thus possible, to select for maintenance of both plasmids in *E. coli* grown in medium containing both antibiotics (ampicillin and kanamycin). The transcription of the T7 RNA polymerase gene is inducible by heat shock. This is due to the fact that the gene is cloned behind a λ P_L promotor and is dependent for expression on a heat sensitive λ repressor encoded on the plasmid pGP1-7 (Tabor and Richardson 1985). Cultures can therefore be grown at low temperature (30°C) to the required cell density. Transcription of T7 RNA polymerase is then induced by heat shock (42°C) and leads to the transcription of any gene cloned into the multiple cloning site of pT7-7. Addition of rifampicin inhibits the action of endogenous *E. coli* RNA polymerases resulting in the exclusive transcription of the recombinant gene. The host for these experiments is *E. coli* WM1704 which carries the Δ Ion mutation. This mutation is a deletion of the Ion protease that has been implicated in problems with expression of foreign genes in *E. coli*. A deletion mutant should facilitate high level expression.

The second system used cell lines that contain the T7 RNA polymerase gene integrated into the *E. coli* genome on a lysogenic prophage (Studier,

Rosenberg et al. 1990). These cell lines are called BL21(DE3), BL21(DE3)LysS and BL21(DE3)LysE. The gene for the T7 RNA polymerase is under the control of the lacUV5 promoter which is inducible by the addition of IPTG to the growth medium. The gene to be expressed is cloned into the T7 vector and cells containing the pT7-7 construct are grown to the desired cell density when T7 RNA polymerase is induced by addition of IPTG with resulting transcription of the recombinant gene.

There is a problem with both these systems in that there is always some basal level of transcription of the T7 RNA polymerase gene. If the protein to be expressed is highly toxic, then the *E. coli* cells sense this even under non-inducing conditions and the vector is often lost or rearranged. If this occurs, it is better to use the BL21(DE3)LysS and BL21(DE3)LysE strains because these produce lysozyme which inhibits transcription by T7 RNA polymerase at the low level present under non-inducing conditions (Studier and Moffatt 1986).

5.1.1.1 Expression studies with b1-1 cDNA in pT7-7

The initial full length cDNA construct of *b1-1* was prepared assuming that the start site was that proposed by Tymon (Tymon, Kües et al. 1992). This first ATG of *b1-1* is 138 bp upstream of a second in frame ATG which was later shown to be sufficient for function in transformation assays (Asante-Owusu 1993). The only cDNA clone available for *b1-1* (pb1-1cDNA) lacked 133 bp of the 5' terminus equivalent to the DNA sequence between the two in-frame ATGs. Fortunately, this region does not contain any introns. It could be excised from the genomic clone pAMT1 and fused to the cDNA clone pb1-1cDNA to yield the full length cDNA clone pEb1-1 (origin and position of pAMT1 and pb1-1cDNA is shown in figure 2.1). The cloning

strategy is shown in figure 5.1. Figure 5.2 shows an ethidium stained agarose gel with various digests illustrating the cloning strategy.

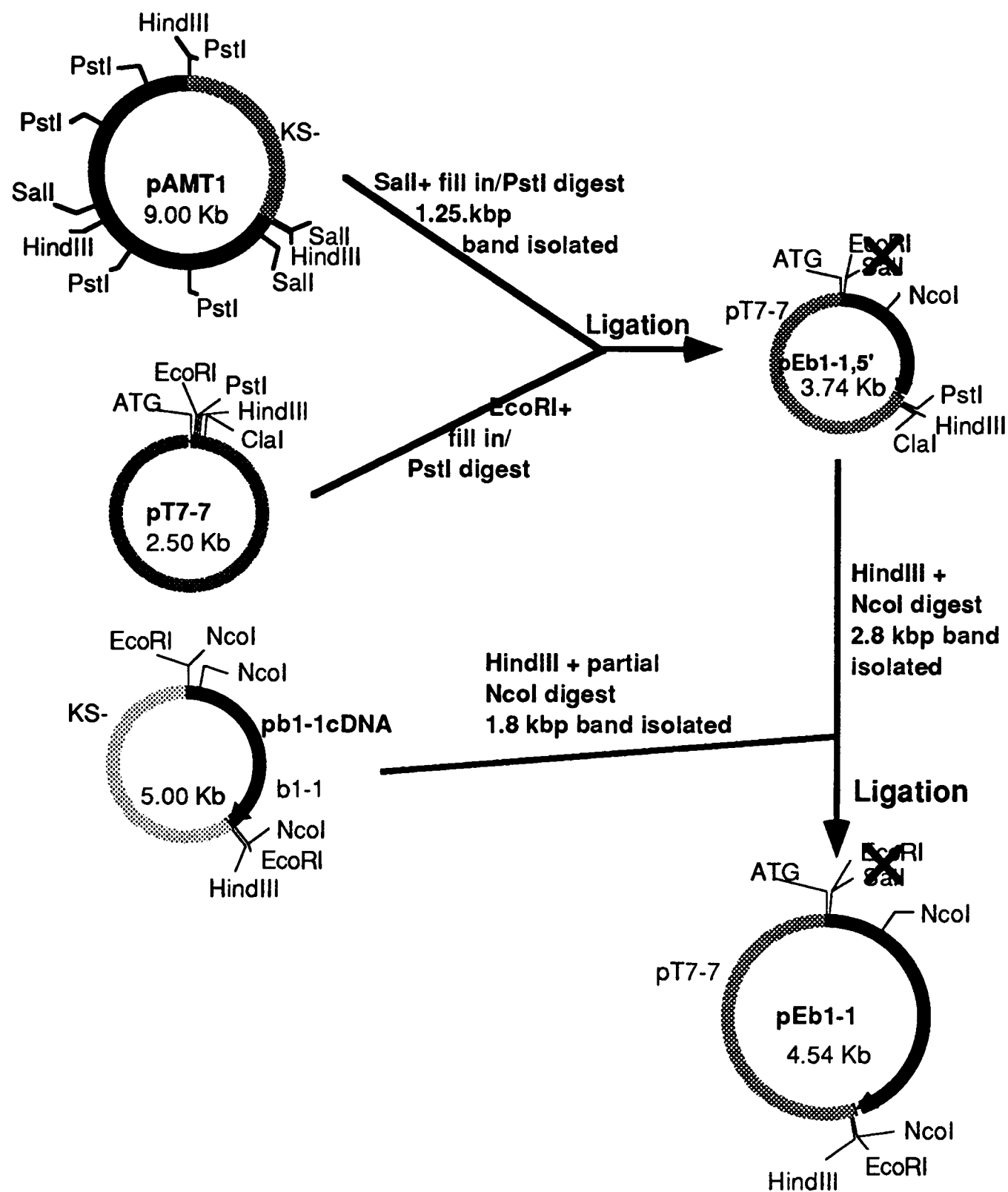


Figure 5.1: Cloning strategy for the construction of pEb1-1. Note that the genomic insert cloned in pAMT1 was a doublet. Restriction sites marked with an X were lost due to in-filling reactions

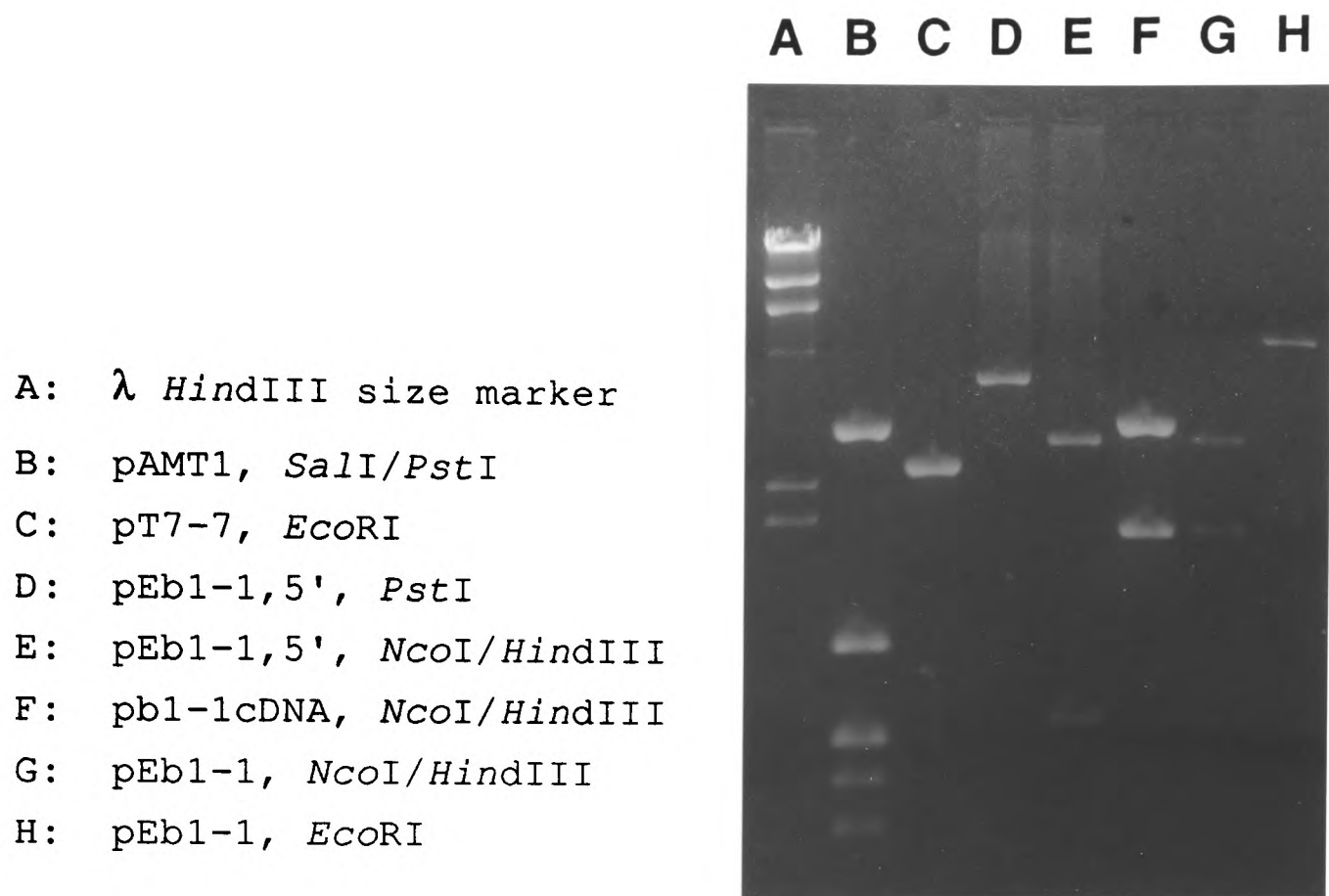


Figure 5.2: Photograph of an ethidium stained agarose gel showing restriction digest patterns of clones used in the preparation of pEb1-1

The cloning strategy necessitated the addition of a few codons 5' to the 1st ATG so that transcription of pEb1-1 starts at the ATG provided in the pT7-7 vector. The sequence of the 5' fusion point is shown in figure 5.3.

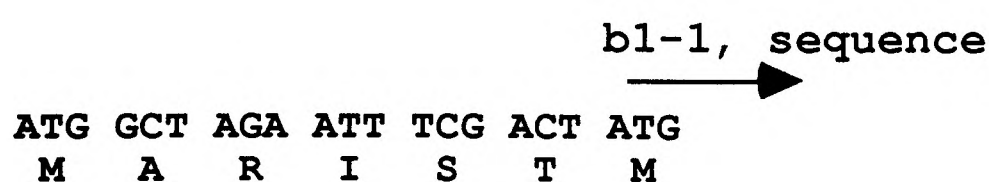


Figure 5.3: Fusion point of the *b1-1* gene starting at the first ATG in pEb1-1

pEb1-1 was used to transform a compatible strain of *C. cinereus* (LT2) and shown to promote clamp cell development. This result was interesting for a number of reasons. Firstly, it showed that the rather complex cloning strategy yielded a functional gene. Secondly, it showed, that the introns normally present in the *C. cinereus* gene do not have an essential function

for gene expression in *C. cinereus*. Thirdly, if the gene starts at this first ATG, then no promotor sequences 5' to the ATG are necessary to allow transcription at a level sufficient for sexual development in *C. cinereus*. Fourthly, it may be noted that the cDNA of *b1-1* differs from the genomic sequence at one position in the homeodomain region, probably due to a misincorporation during the synthesis of the cDNA. The T at position 773 of the genomic sequence (numbering according to (Tymon, Kües et al. 1992)) is replaced by a C in the cDNA sequence leading to an amino acid change from W to R near the end of the first helix of the homeodomain (position 21 of the homeodomain according to the homeodomain numbering conventions used in chapter 4). This amino acid replacement clearly does not abolish the function of the *b1-1* protein in transformation assays, which means that amino acid changes in this region of the homeodomain can be tolerated.

Two 3' truncations of the *b1-1* gene cloned in pEb1-1 were prepared. pEb1-1Nco was obtained by digesting pEb1-1 DNA with *Nco*I. The resulting 2.8 kbp band was gel-purified and religated. pEb1-1Cla was prepared similarly by digesting pEb1-1 with *Cla*I and isolating the 3.25 kb band. pEb1-1Nco contains the first 300 nucleotides, pEb1-1Cla the first 700 nucleotides of the *b1-1* gene (counting from the first ATG; the homeodomain then extends from nucleotides 500 to 720).

All three plasmids, pEb1-1, pEb1-1Nco and pEb1-1Cla were tested in an expression experiment using the two plasmid system. For each plasmid several independently isolated clones were tested but none overexpressed the *b1-1* protein following induction (data not shown). pEb1-1 and pEb1-1Cla were also tested for overexpression of *b1-1* using the three *E. coli* strains lysogenic for the T7 RNA polymerase (BL21(DE3), BL21(DE3)LysS and BL21(DE3)LysE). Again no overexpression of *b1-1* protein was observed

(data not shown). These results suggested that the b1-1 protein expressed in these systems may be too toxic to *E. coli* and is either degraded very rapidly, or the plasmid carrying the gene is lost or mutated. Similar problems have been reported in studies with other homeodomain proteins which could only be poorly expressed using a T7 expression system (Budd and Jackson 1991).

5.1.2. Expression studies with the 5' end of *b2-1* cloned into pT7-7

In a second set of experiments, the first 402 bp of the *b2-1* gene from A42 were tested for overexpression in *E. coli* using the pT7-7 vector system. The cloning strategy is illustrated in figure 5.4. This sequence does not contain any introns and should be translated correctly by *E. coli*.

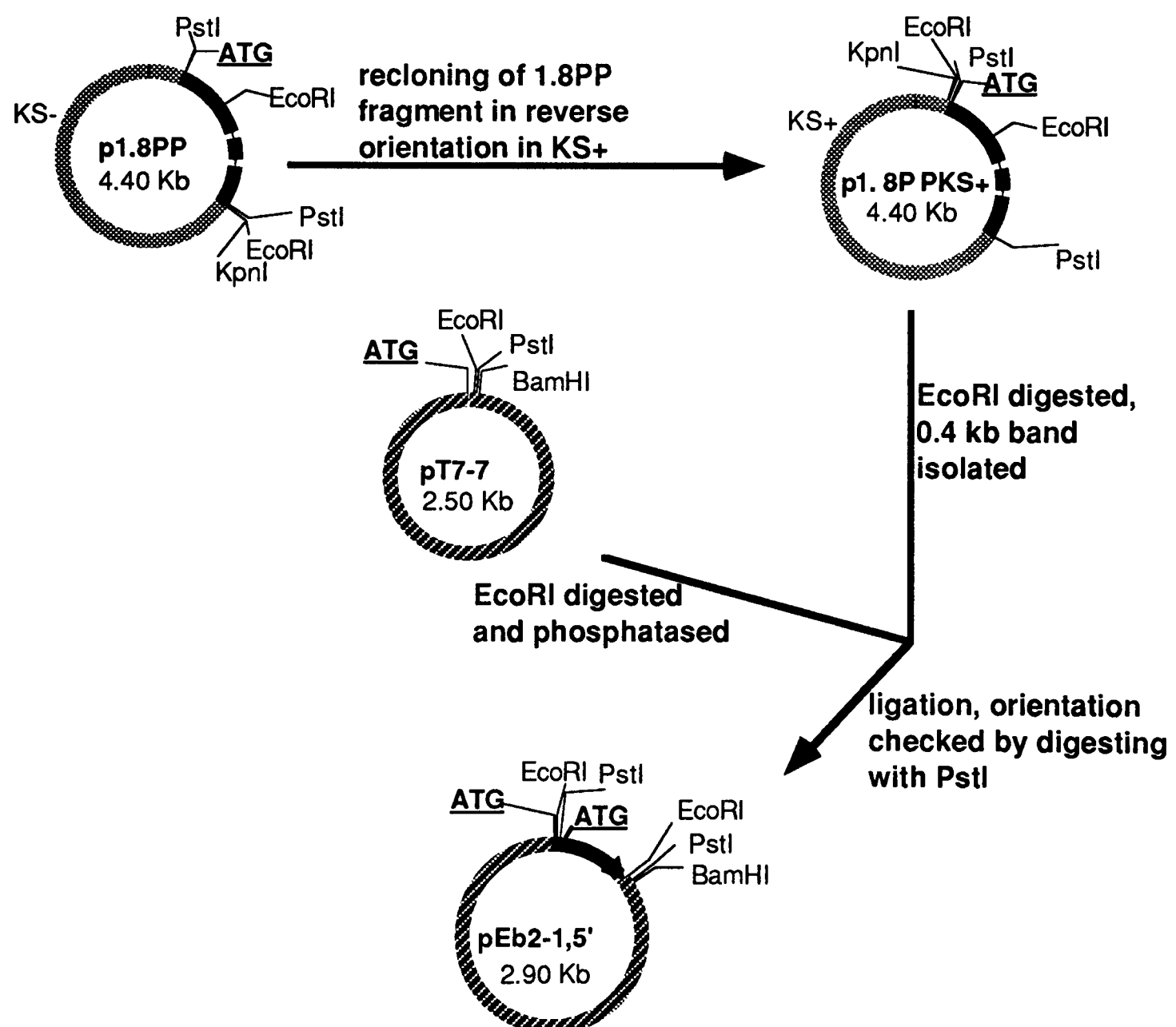


Figure 5.4: Cloning strategy for the construction of the pEb2-1,5' clone. p1.8PP contained a 1.4 kbp PstI fragment of genomic DNA from *b2-1* of A42

The clone was designated pEb2-1,5' and figure 5.5 shows an ethidium stained agarose gel with various restriction digests illustrating the cloning strategy used.

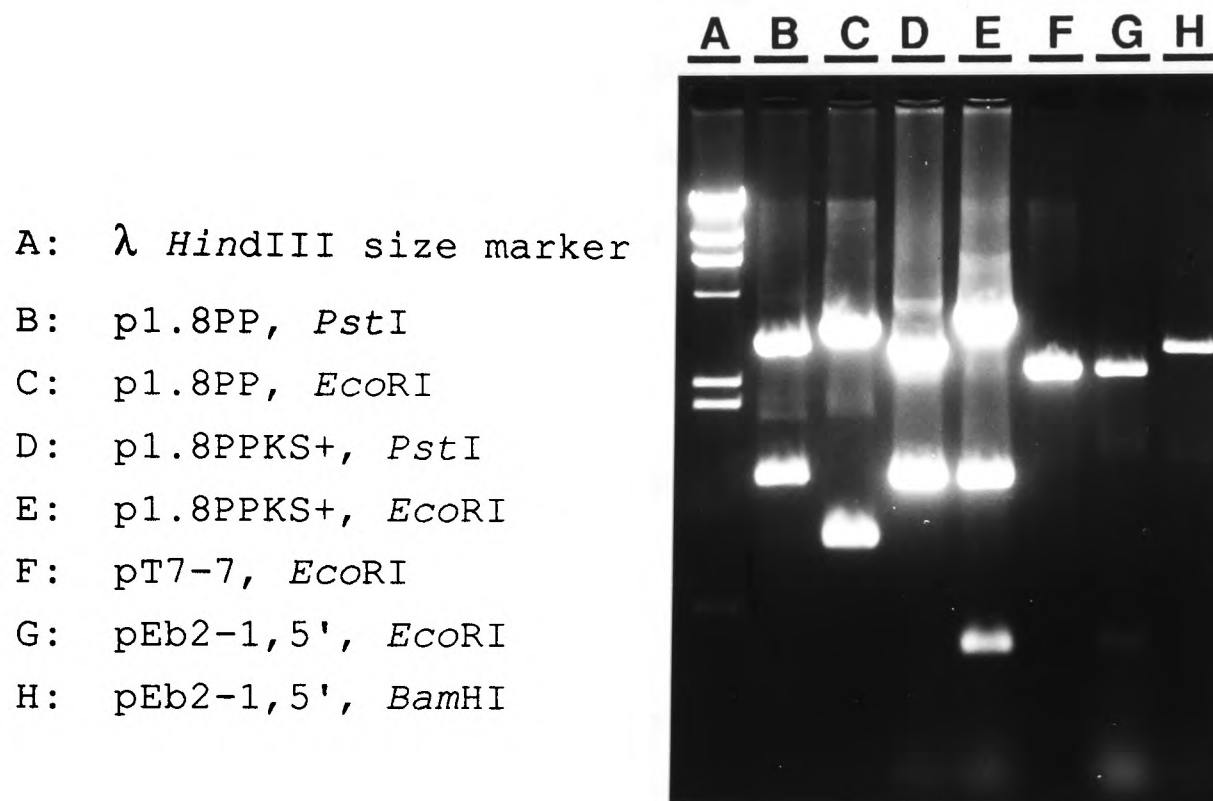


Figure 5.5: Ethidium stained agarose gel showing restriction digest patterns of plasmids used in the construction of pEb2-1,5'

The fusion point of the 5' end of the *b2-1* gene to the pT7-7 multiple cloning site in pEb2-1,5' is shown in figure 5.6.

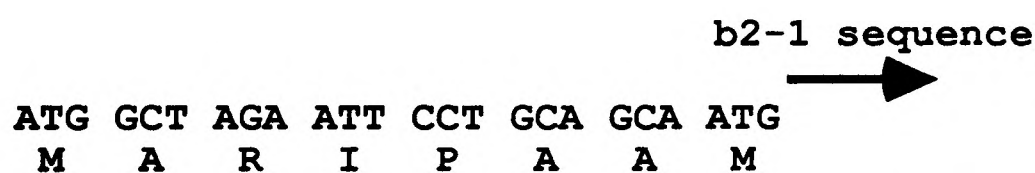


Figure 5.6: Fusion point of the *b2-1* gene in pEb2-1,5'

pEb2-1,5' was tested for expression using the two-plasmid system in the host WM1704 and the T7 expression system with the three lysogenic BL21 strains. Neither system yielded any detectable b2-1 protein.

The T7 overexpression approach was abandoned at this point since it failed to yield either of the two A42 b proteins at levels that could be detected on Coomassie stained SDS gels of crude *E. coli* lysates. A straightforward T7 system is only useful for preparing protein to raise antisera if the recombinant protein is produced at very high levels. The basic T7 system used here does not allow easy purification of the recombinant protein so a multi-step purification protocol would have to be developed which would be impossible without a functional assay for the protein of interest.

5.2 Plasmid constructs for overexpression of the A42 b genes in *E. coli* using the pGEX vector system

In the pGEX vector system, the gene of interest is fused to part of the glutathione S-transferase gene of *Schistosoma japonicum* and expressed as a fusion protein. The glutathione S-transferase (GST) moiety of the fusion protein forms an affinity tail for purification using glutathione sepharose (Smith and Johnson 1988). The pure fusion protein can then be cleaved to separate the GST and the recombinant protein using the proteases thrombin (pGEX2T) or factorXa (pGEX3X).

5.2.1 Preparation of a construct containing the 5' half of the *b1-1* gene

The 5' half of the full length cDNA of *b1-1* in pEb1-1 (starting at the 1st ATG) was fused to the GST gene in the expression vector pGEX3X to yield the expression construct pHDX3. The cloning strategy is illustrated in figure 5.7.

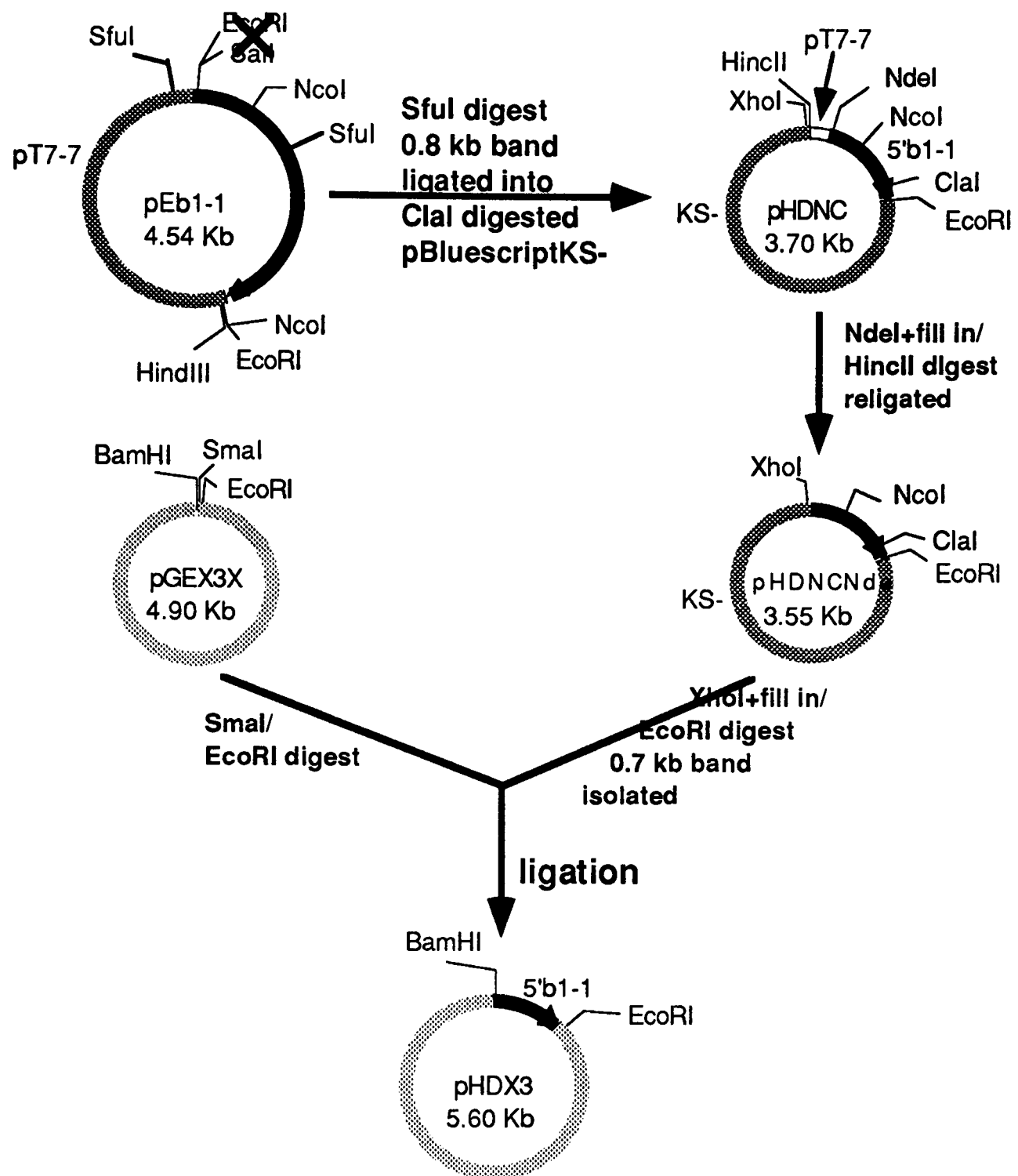


Figure 5.7: cloning strategy for the construction of pHDX3. Restriction sites marked with an **X** were lost due to in-filling reactions. The clone pHDNC contained parts of the pT7-7 vector sequence that had to be removed prior to the construction of the pHDX3 fusion construct

Figure 5.8 shows an ethidium stained agarose gel with restriction digests of clones illustrating the cloning strategy.

- A: λ HindIII size marker
 B: 0.8 kbp SfuI band from pEb1-1
 C: pBluescript, EcoRI
 D: pHDNC, EcoRI/XhoI
 E: pHDNC, XhoI
 F: pHDNCNde, XhoI
 G: pHDNCNde, EcoRI/XhoI
 H: pGEX3X, EcoRI
 I: pHDX3, BamHI/EcoRI

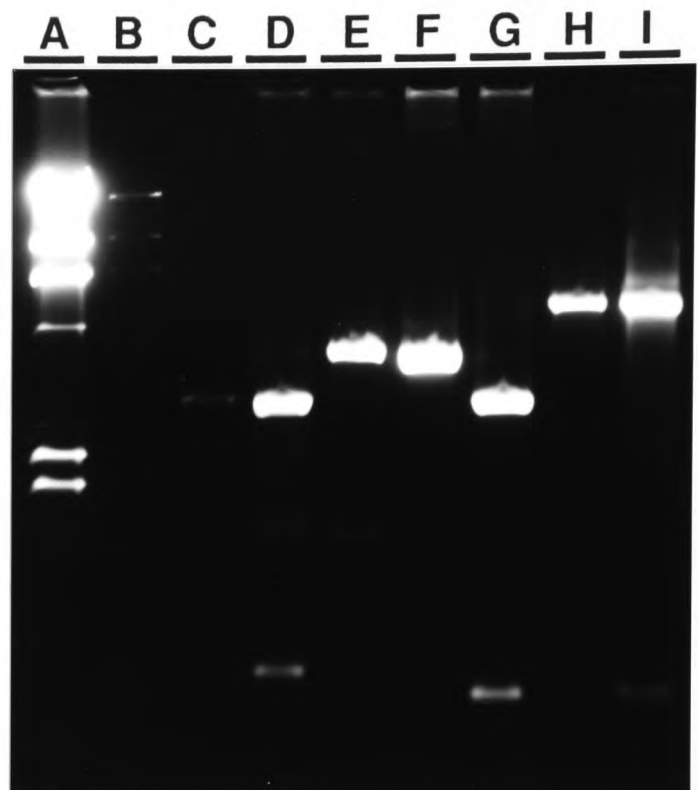


Figure 5.8: Ethidium stained agarose gel showing restriction digest patterns of plasmids used in the construction of pHDX3

Figure 5.9 shows the fusion point of the *b1-1* sequence (originating from pEb1-1) and the GST gene sequence in pHDX3

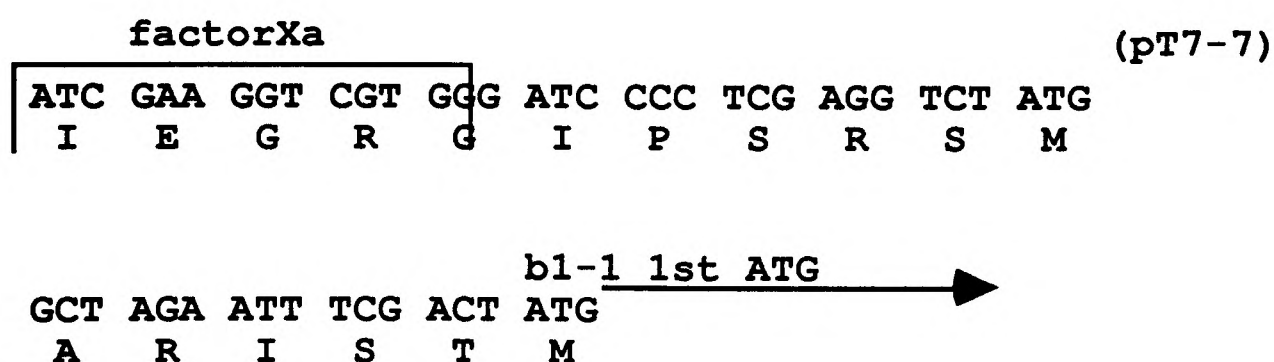
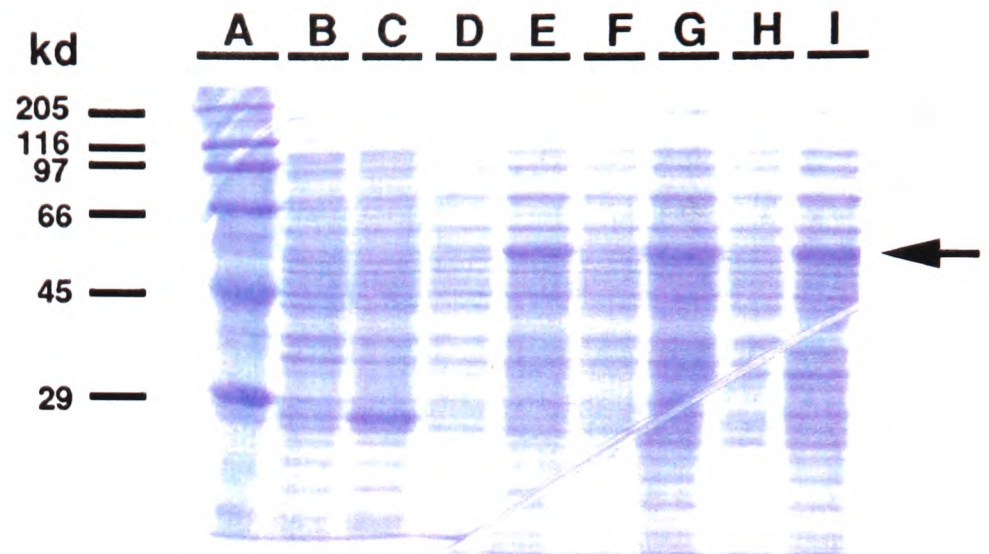


Figure 5.9: Fusion point of the GST gene and the *b1-1* sequence in pHDX3 (FactorXa cleavage site and ATG originating from pT7-7 as indicated)

Several independently isolated clones containing this plasmid were tested for overexpression in *E. coli*. All showed fairly strong expression of a novel protein after induction with IPTG as shown in figure 5.10 (a). The novel protein has the expected size for a fusion protein composed of GST and the first 212 amino acids of the b1-1 protein.

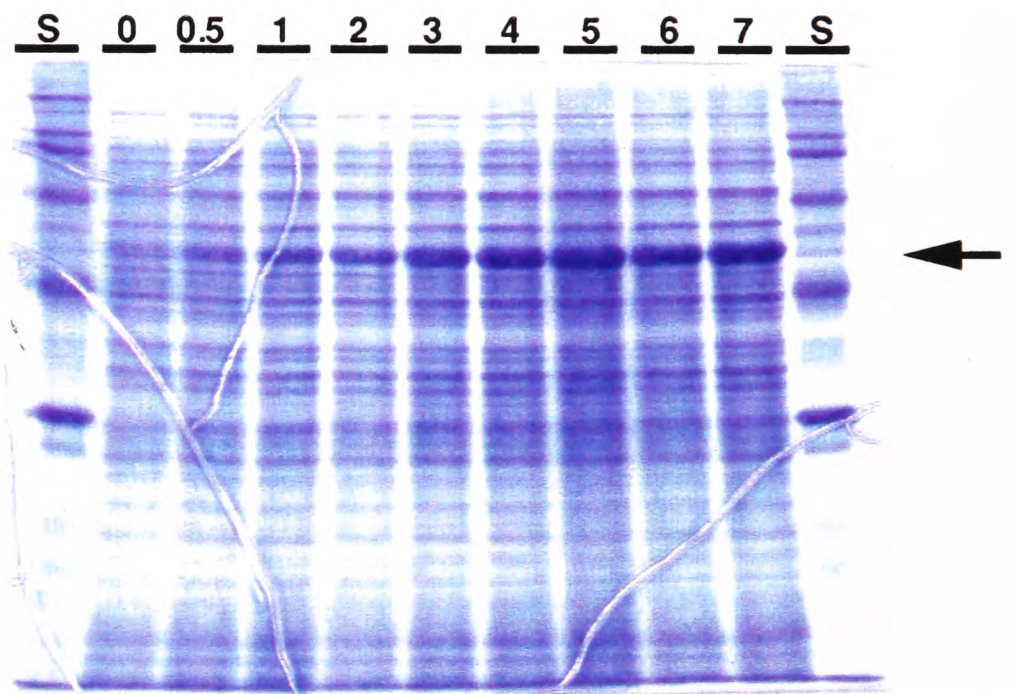
(a)

A: high molecular weight marker
B: pGEX control, uninduced
C: pGEX control, induced
D/F/H: pHDX3 clones, uninduced
E/G/I: pHDX3 clones, induced



(b)

S: high molecular weight marker
0-7: samples of *E. coli* carrying the pHDX3 plasmid after 0 to 7 hours after induction with IPTG



(c)

S: high molecular weight marker
A: uninduced sample
B: induced sample
C: soluble fraction after sonication
D: insoluble fraction after sonication

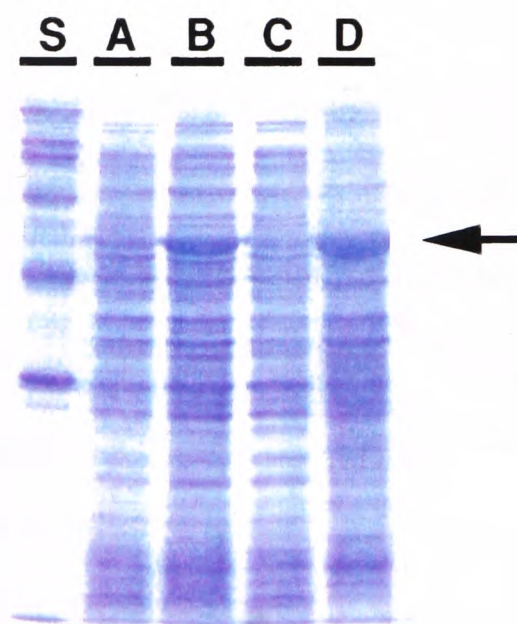


Figure 5.10: (a) production of a novel protein after induction with IPTG in *E. coli* cells carrying the pHDX3 plasmid. (b) time course of the expression of the HDX3 fusion protein. (c) samples taken from a large scale preparation of the HDX3 fusion protein proving that the recombinant protein was insoluble (arrow = position of recombinant protein).

Figure 5.10 (b) illustrates a time course for expression of the HDX3 fusion protein. It shows that maximum levels are obtained after 4 to 5 hours induction with IPTG. Microscopic examination of the *E. coli* cells after induction showed that the foreign protein was being sequestered in inclusion bodies, insoluble deposits of protein that allow *E. coli* to tolerate such high levels of foreign protein. *E. coli* adopts an elongated cell shape with inclusion bodies at either end of the rod like cell. The presence of inclusion bodies means that the desired protein is insoluble and this has implications for any purification attempts.

The HDX3 protein was then expressed and purified on a large scale as described in section 2.15.2. Figure 5.10 (c) shows various samples taken during the course of this experiment, confirming that the recombinant HDX3 protein was exclusively in the insoluble fraction. The insoluble protein was now solubilised in order to allow affinity purification using glutathione agarose. However, the solubilised protein did not fully renature and did not bind to the affinity matrix glutathione agarose (data not shown).

It was decided at this stage not to pursue the isolation of the HDX3 protein from the pHDX3 construct. Sequencing of the *b1-3* gene (chapter 4) and *in vitro* mutagenesis of the first ATG (Asante-Owusu 1993) suggested that the second in frame ATG in *b1-1* is the actual start codon. This meant that the HDX3 protein contained some 49 additional amino acids that would not be present in the *b1-1* protein synthesised in *C. cinereus*. It would be especially undesirable to use the HDX3 protein to raise antibodies since antibodies raised to epitopes in these additional 49 amino acids would not react with *b1-1* produced *in vivo*, but might give non-specific side reactions.

5.2.2 New plasmid

The cDNA clone, pb (Kües et al. 1992) s codon, allowing the new expression cor expression of the 5' illustrated in figure 162 amino acids, clo to the GST protein i

Figure 5.11 shows a trating the cloning pXCT.

- A: λ HindIII si
- B: pGEX3X, EcoR
- C: pb1-1cDNA, B
- D: pb1-1X3, Bam
- E: pb1-1T2, Bam
- F: pb1EXT, BamH
- G: pGEX2T, PstI
- H: pXBT, PstI
- I: pXCT, PstI

Figure 5.11: Ethid patterns of plasmids

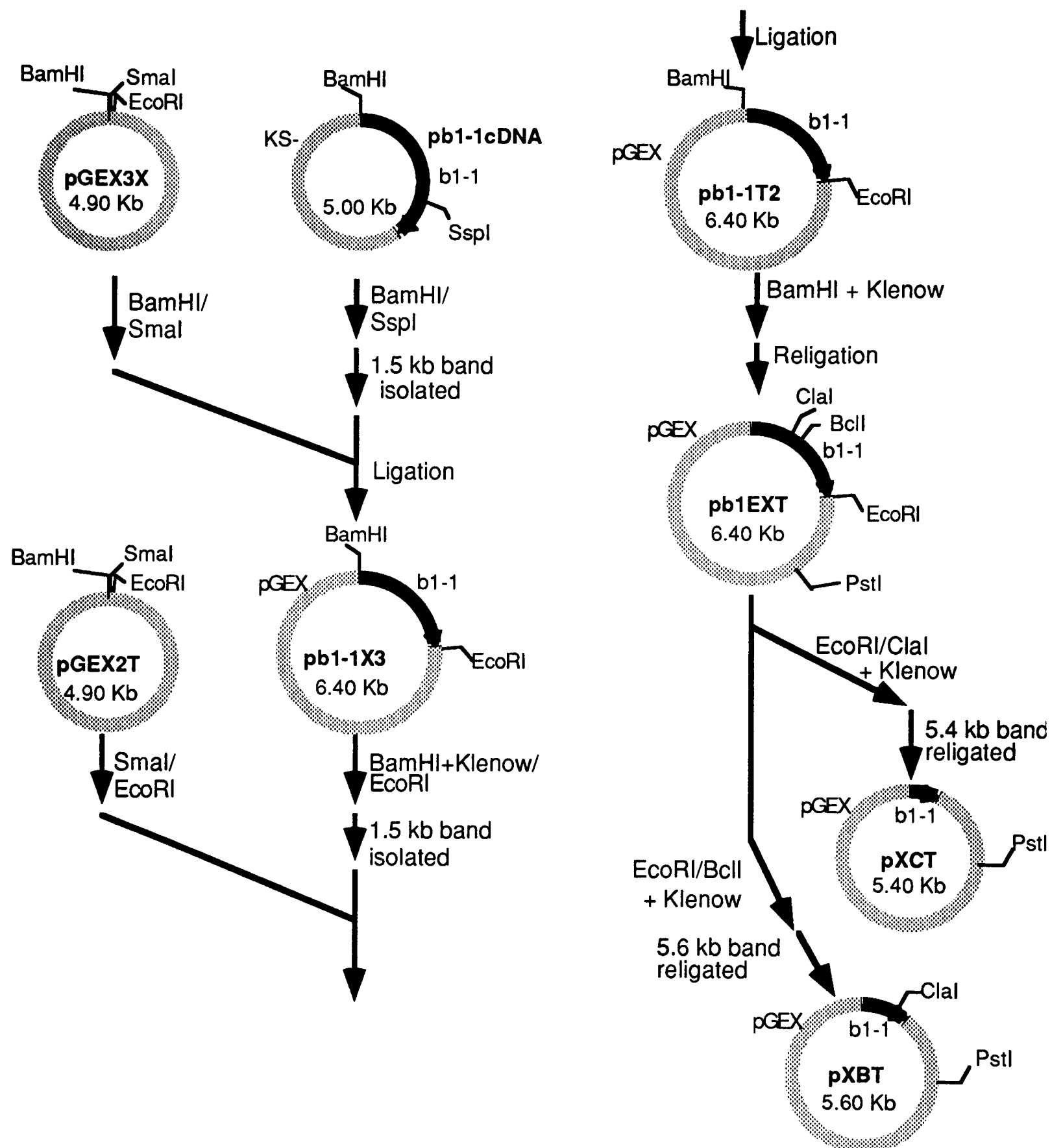


Figure 5.12: Cloning strategy for the construction of pXBT and pXCT.

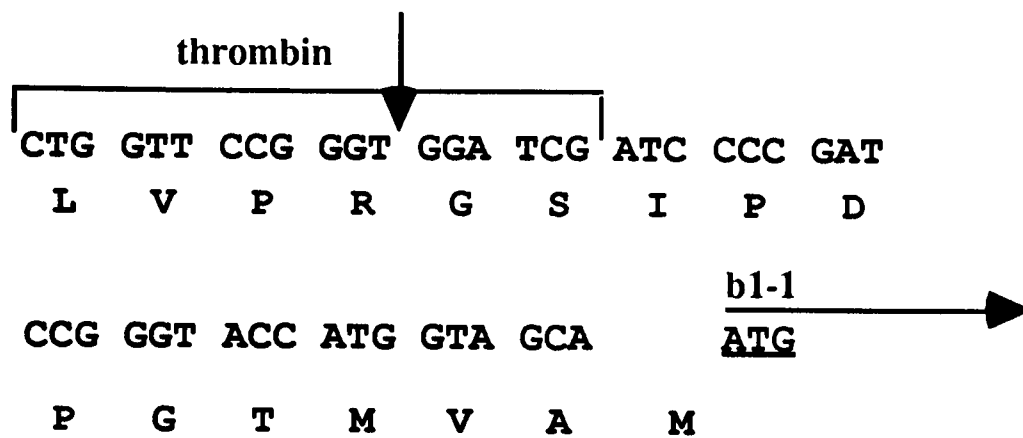


Figure 5.13: In-frame fusion region of the pXBT and the pXCT constructs at the 5' border to the glutathione S-transferase gene.

Various independently isolated clones containing each plasmid were used in small scale expression experiments and in all cases the appearance of a new protein band was seen after induction with IPTG (data not shown). These proteins were of the expected size and designated XB1 and XC1 protein respectively. The time course of the expression of both proteins was monitored and for both proteins, optimum levels of expression were observed after 4 hours induction. Microscopic examination of the cells following induction with IPTG confirmed the presence of inclusion bodies indicating that both the XB1 and the XC1 protein were insoluble.

A large scale inclusion body preparation was carried out to produce XB1 and XC1 protein for injection into rabbits in order to raise polyclonal antibodies. SDS-PAGE analysis of an aliquot of the inclusion bodies showed that neither protein preparation was pure enough to be used directly for raising antibodies. The XB1 protein was further purified by preparative SDS-PAGE. Figure 5.14 (a) shows an SDS-PAGE that summarises the purification procedure used to obtain 10 mg of pure XB1 protein that was subsequently used to immunise two rabbits.

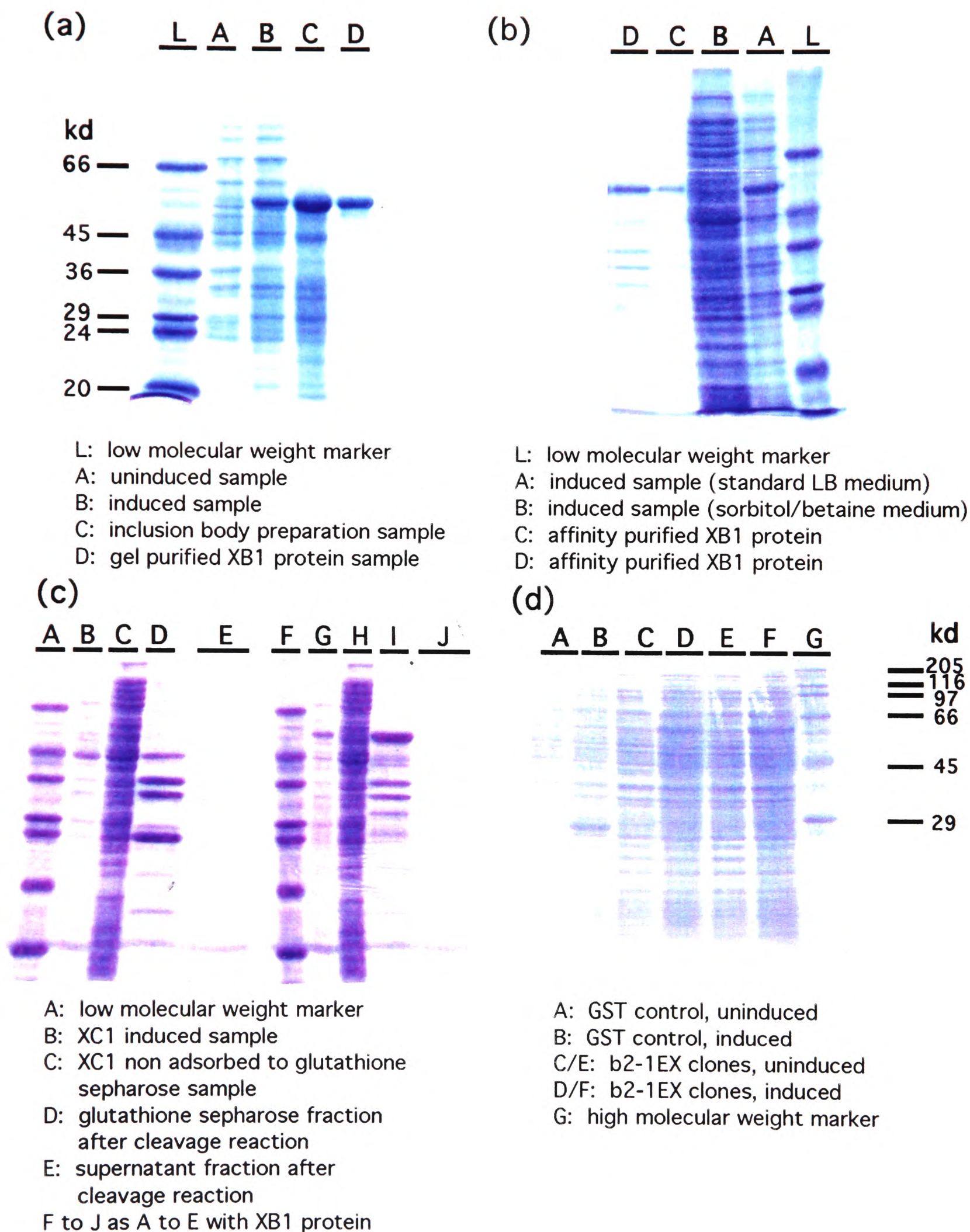


Figure 5.14: (a) samples taken during the purification of the XB1 fusion protein. (b) results of glutathione affinity purification of the XB1 protein expressed in the presence of sorbitol and betaine. (c) results of a thrombin cleavage experiment using the XB1 and XC1 fusion proteins. (d) production of a novel protein after induction of *E. coli* cells harbouring the pb2-1EX plasmid.

Both proteins were subsequently expressed in the presence of 0.6 M sorbitol and 2.5 mM betaine in the growth medium (section 2.15.5) which results in the production of partly soluble protein. The Xb1 and XC1 proteins could then be purified from the respective crude protein extracts using the affinity matrix glutathione sepharose as shown in figure 5.14 (b). An attempt to cleave these purified fusion proteins with thrombin is shown in figure 5.14 (c). Only the XC1 protein was cleaved under the conditions used. However, after cleavage, the b1-1 part of the XC1 fusion protein disappeared, i.e. only the GST band is still visible after electrophoresis. This suggests that the part of the b1-1 protein expressed in XC1 must be stabilised in the GST-fusion protein and is prone to proteolytic degradation after cleavage.

It is not clear why the XB1 fusion protein was not cleaved by thrombin under the same conditions that readily cleaved the XC1 fusion protein. pXBT and pXCT were derived from exactly the same initial clone, pb1EXT by 3' truncations of the *b1-1* sequence (see figure 5.11). Thus the thrombin recognition site is identical in pXBT and pXCT. It can only be suggested that the longer XB1 protein might fold back and obstruct the thrombin cleavage site.

5.3 Plasmid construct for overexpressing *b2-1* gene sequence using the pGEX vector system

The *b2-1* cDNA clone pb2-1cDNA contains an almost full length copy of this gene lacking 173 bp at the 5' end (the position of pb2-1cDNA in relation to the A42 factor is shown in figure 2.1). A 1 kbp *Bam*HI/*Hinc*II fragment of the 5' end of this clone was cloned into the expression vector pGEX3X as illustrated in figure 5.15. The predicted protein should lack the first N-

terminal 58 amino acids. The purpose of this experiment was to test whether parts of the *b2-1* gene could be expressed as GST fusion proteins after attempts to express *b2-1* using the T7 expression systems failed (section 5.1.2).

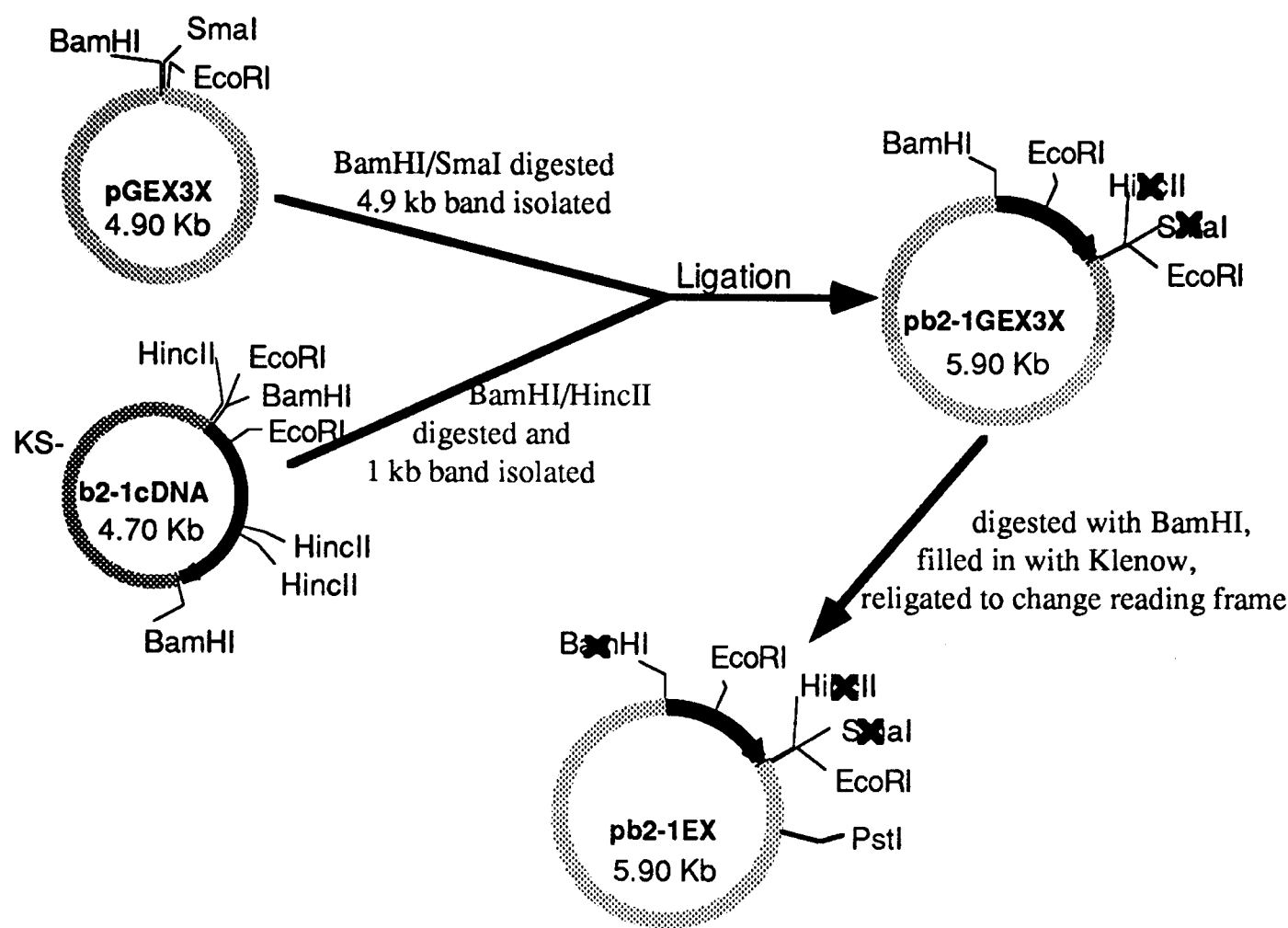


Figure 5.15: Cloning strategy for the construction of the pb2-1EX expression vector. Restriction sites marked with an **X** were lost due to in-filling reactions.

The fusion point of the *b2-1* coding region to the GST gene of pGEX3X is illustrated in figure 5.16.

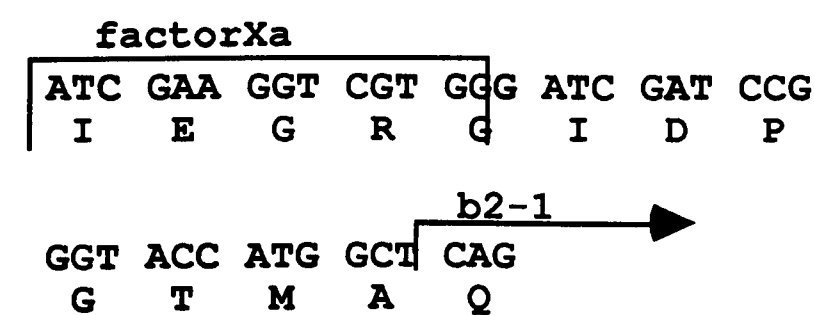


Figure 5.16: Fusion of the *b2-1* coding region to the GST gene of pGEX3X (factorXa recognition site as indicated)

Figure 5.17 shows an ethidium stained agarose gel with various restriction digests illustrating the cloning strategy used in the preparation of pb2-1EX. Several independently isolated clones of *E. coli* XLIBLue carrying the pb2-1EX plasmid were subjected to small scale expression experiments to check for the production of a recombinant protein. Extracts of all clones tested revealed an additional band with the predicted size for the b2-1EX fusion protein (example shown in figure 5.14 (d)).

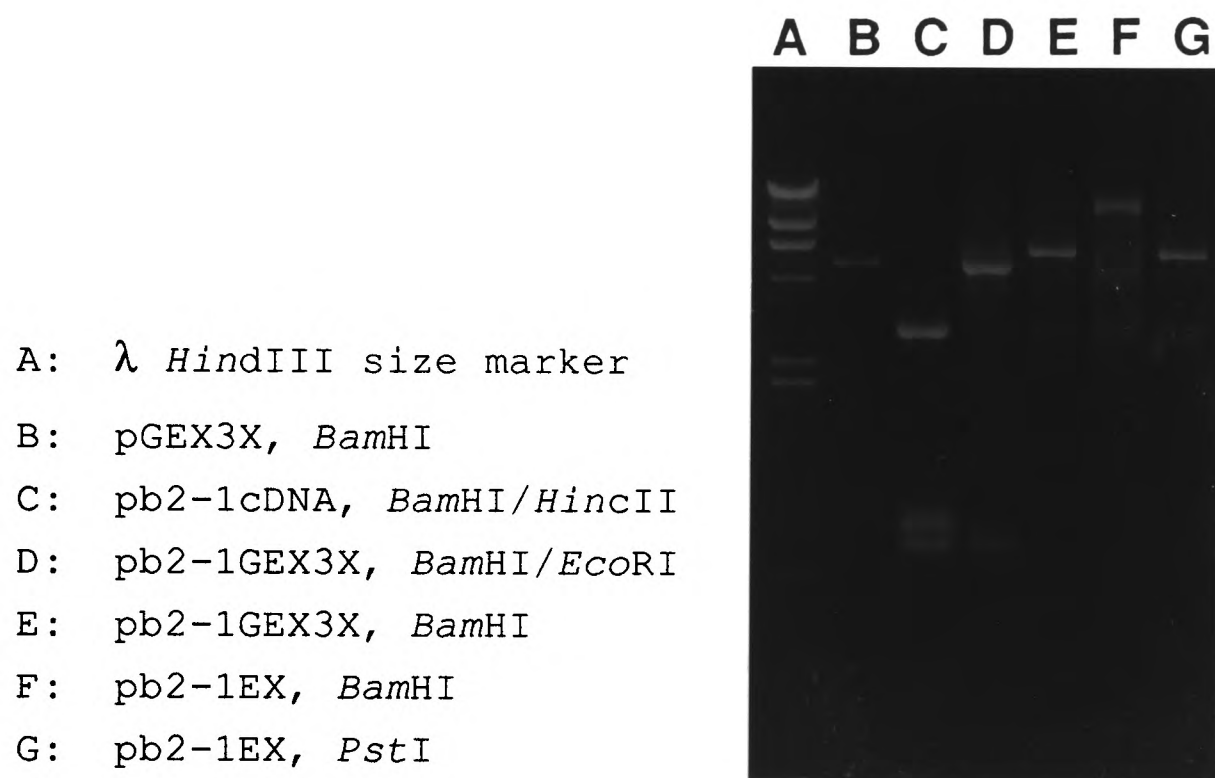


Figure 5.17: Ethidium stained agarose gel showing restriction digest patterns of clones used in the preparation of pb2-1EX

Thus as for the b1-1 protein, fusion to GST proves to be a successful strategy for overexpression even though attempts to express both genes in T7 vectors failed. Since the XB1 protein (containing parts of b1-1) was overexpressed to much higher levels than b2-1EX, no attempt was made to raise antibodies to b2-1.

5.4 Immunisation of two rabbits with the XB1 fusion protein

Pre-immune sera were collected from two rabbits identified by the numbers 4922 and 4923. The pre-immune sera were tested by Western blotting for non-specific reaction to the XB1 protein in crude *E. coli* extract and crude *C. cinereus* protein extract. Both pre-immune sera showed no reaction with the XB1 fusion protein. There was some moderate reactivity to the *E. coli* crude cell lysate and some reactivity to the crude *C. cinereus* protein extract. It was assumed that after immunisation the antisera could be used at far lower dilutions than the one used for testing the pre-immune sera (1:100), and that this would effectively reduce the non-specific background reaction with *C. cinereus* crude protein extract. Both rabbits were used to raise polyclonal antibodies against the XB1 fusion protein.

Approximately 1 mg of inclusion body preparation was gel purified for each immunisation. Assuming an 80% recovery, approximately 800 µg of XB1 protein were used for each immunisation (see table 5.1). Injection of macerated gel slices has proved to be a very effective way of raising antibodies to pGEX fusion proteins (Oettinger, Pasqualini et al. 1992).

Table 5.1: Immunisation programme for the two rabbits 4922 and 4923

1 st pre-bleed	21-4-93
2 nd pre-bleed	26-4-93
1 st immunisation	28-4-93
2 nd immunisation	26-5-93
1 st bleed	1-6-93
3 rd immunisation	17-6-93
2 nd bleed	22-6-93
4 th immunisation	5-8-93
3 rd bleed	17-8-93
5 th immunisation	20-10-93
final bleed	27-10-93

All immunisations were carried out by adding an equal volume of Freund's adjuvant to the macerated gel slice and directly injecting this mixture into the rabbits 4922 and 4923. Antisera were prepared according to standard procedures and aliquots were stored at -20°C.

5.5 Analysis of antisera

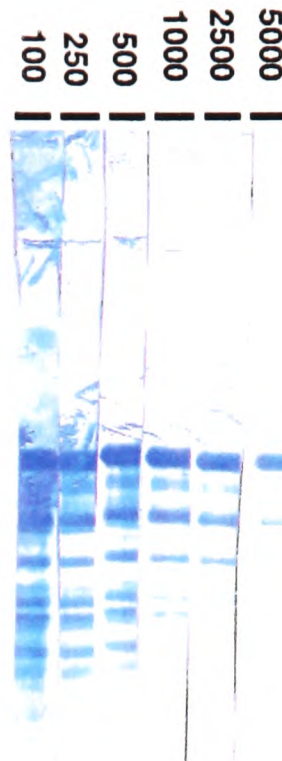
5.5.1 Analysis of antisera by Western blotting

5.5.1.1 Analysis of antisera by western blotting using the XB1 fusion protein

Western blot analysis showed that the antisera taken from both rabbits 4922 and 4923 reacted strongly with the XB1 fusion protein. The 3rd bleed antisera of both rabbits were titred in their reaction with the XB1 fusion protein as shown in figure 5.18. It is apparent that the antisera react with the XB1 protein even at a dilution of 1 : 5,000. Moreover the background was much reduced at lower dilutions of the primary antibody. This experiment did not distinguish whether the antibodies reacted with the *C. cinereus* b1-1 portion of the XB1 fusion protein or merely with the GST component. This problem could not be solved using the XB1 fusion protein, because as discussed earlier (section 5.2.2), the soluble form of XB1 could not be cleaved by thrombin to separate the b1-1 sequence from the GST portion.

(a)

numbers indicate
the dilution factor
used for antisera (3rd
bleed) of the rabbit
identified by the
number 4922



(b)

R: rainbow [™] molecular
weight marker
I: 0.1 μ l aliquot
of XB1 inclusion
body preparation

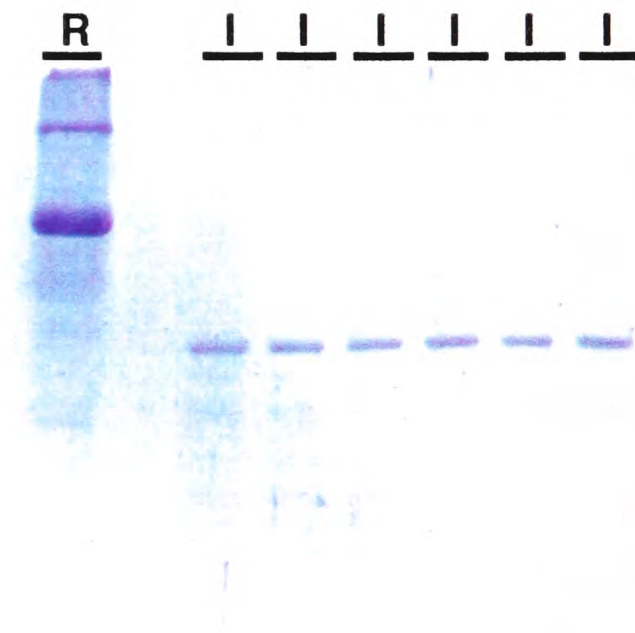


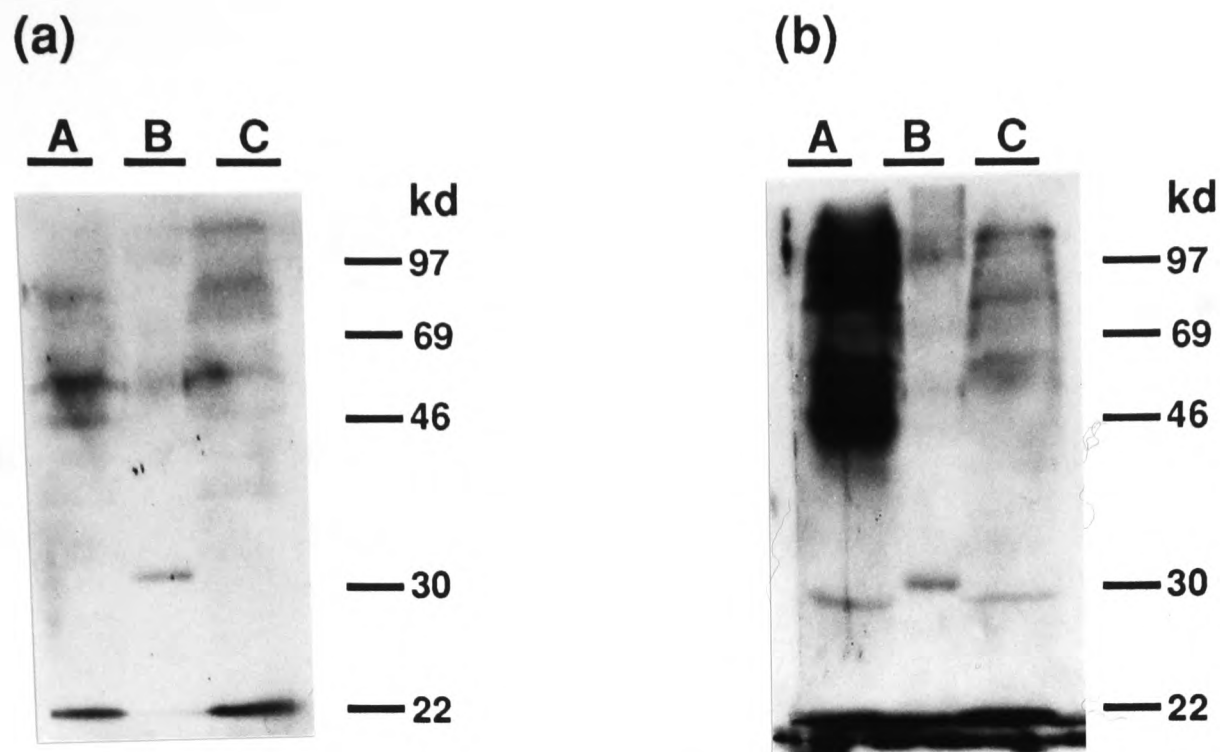
Figure 5.18: (a) dilution series of polyclonal antisera in their reaction to the XB1 fusion protein in Western blotting analysis. (b) the same gel stained after Western blotting with Coomassie brilliant blue showing the equal loading of XB1 protein in all tracks

5.5.1.2 Analysis of antisera by Western blotting using *C. cinereus* total protein extracts

The antisera of both rabbits were tested in Western blots with crude total protein extracts of *C. cinereus*. Proteins were extracted from the A42

monokaryon JV6, the dikaryon produced by mating of JV6 to the A6 monokaryon H9 (A42 x A6) and a mutant A-null monokaryon NA2, that has no functional A mating type genes (Pardo 1993). It was expected that a specific reaction to the b1-1 protein would not be observed in the protein extracts of NA2, but would be expected for extracts of both the monokaryon and the dikaryon. For this experiment, the alkaline phosphatase colour reaction detection procedure failed to detect any convincing positive signals.

The chemiluminescence Western blot detection procedure (Amersham, ECL-Western blot detection kit) is considered to be more sensitive. Using this technique, the background already detected in the alkaline phosphatase colour reaction was even more pronounced. No specific bands for the b1-1 protein could be detected as shown in figure 5.19 (a). Although there is some reactivity to a novel band of approximately 30 kd after immunisation this band appears in lanes corresponding to protein extracts from JV6 and the NA2 null mutant. It could represent cross-reactivity of the antisera raised to some *C. cinereus* protein, maybe a glutathione S-transferase, as the XB1 protein used for immunisation contained part of a bacterial glutathione S-transferase protein. The experiment was repeated using pre-cleared antisera but the background problem persisted. mRNAs for the *C. cinereus* mating type proteins are very rare and with a predicted short half life, the concentration of the proteins can be assumed to be equally low. This is likely to make the detection of b1-1 by Western blot analysis in crude extracts very difficult.



4923 pre-immune serum

4923 3rd bleed antiserum

Figure 5.19: Western blot analysis of *C. cinereus* NA2 (A) and JV6 (C) crude protein extract (RainbowTM marker in tracks (B); detection using the ECL chemiluminescence system (Amersham))

5.5.2 Analysis of antisera by immunoprecipitation of *in vitro* translated b1-1 protein

The *b1-1* cDNA clone pb1-1cDNA was derived from a cDNA library prepared in the lambda vector λ gt10 and had been cloned into the *EcoRI* site of the plasmid vector pBluescript KS-(see figure 2.1 for the position of pb1-1cDNA in relation to the A42 factor). The adapter used in the construction of the cDNA library contained an *NcoI* restriction site which contains a potential ATG start codon. Fortunately the ATG in the *NcoI* recognition site was in-frame with the ATG start codon of the *b1-1* coding sequence (figure 5.20). As shown in figure 5.21, both these ATGs are found immediately downstream of the T7 RNA polymerase promoter in the KS-vector. Hence this plasmid could be used directly for *in vitro* transcription using T7 RNA polymerase.

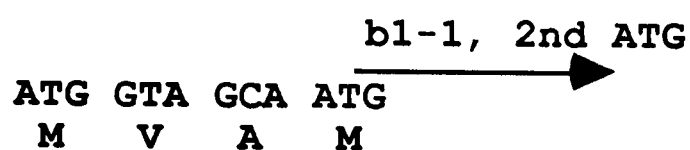


Figure 5.20: 5' region of the *b1-1* cDNA pb1-1cDNA clone showing the two in-frame ATG codons

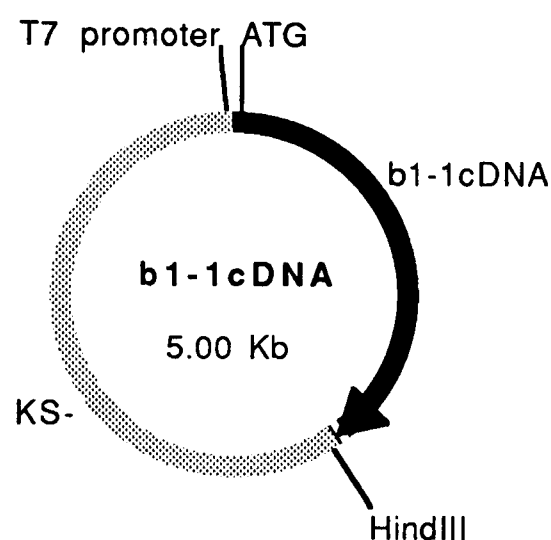
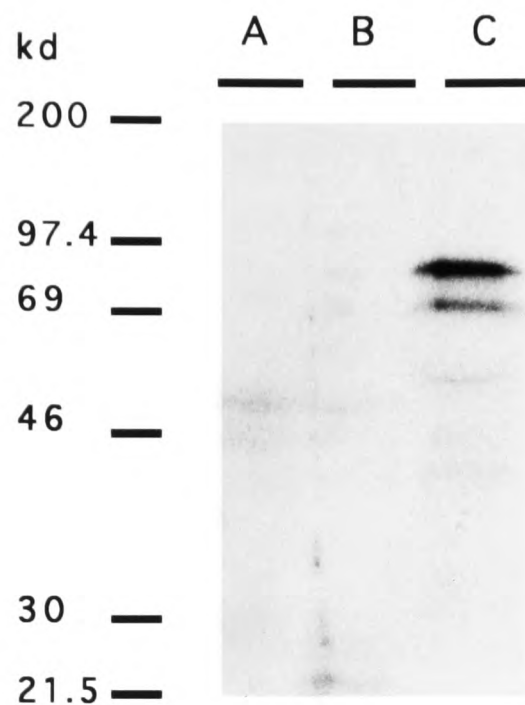


Figure 5.21: Map of the *b1-1* cDNA clone in pBluescript KS- (pb1-1cDNA) showing the location of the T7 promoter upstream of the *b1-1* coding sequence

The *b1-1* protein was produced by *in vitro* transcription/translation and immunoprecipitated with antisera from both rabbits (4922 and 4923). Both samples of antisera specifically immunoprecipitated the *in vitro* translated *b1-1* protein as shown in figure 5.22 for antisera of the rabbit 4923. This result confirmed that the polyclonal antibodies raised against the XB1 fusion protein specifically recognise the *C. cinereus* *b1-1* protein and are thus potentially useful for future studies with this protein.



A: supernatant of immunoprecipitation

B: precipitation with preimmune serum

C: precipitation with b1-1 antiserum

Figure 5.22: Immunoprecipitation of *in vitro* translated b1-1 protein

5.5.3 Analysis of antisera by immunofluorescence

A simple test of whether the antibodies could detect b1-1 *in situ* in cells of *C. cinereus* was attempted. Mycelia of the A42 monokaryon (JV6) and of the A-null mutant monokaryon NA2 was fixed and tested for immunofluorescence with pre-cleared pre-immune and various dilutions of antisera (1:100/ 1:1,000/ 1:5,000). However, no specific staining could be detected even though the background fluorescence was low (data not shown). This may be another reflection of the low intracellular concentration of the *C. cinereus* A mating type proteins.

Chapter 6

Analysis of protein-protein interactions of the *C. cinereus* A mating type proteins

The model that best explains a compatible A mating type interaction is illustrated in figure 6.1. It shows the *b* gene-pairs of the A42 and A6 mating type loci. The different allelic forms of these genes are responsible for compatibility of A42 and A6 (see chapter 1). Following mating, cells have all four genes. A compatible interaction is between *b1-1* and *b2-3* and between *b1-3* and *b2-1*. This non-allelic gene interaction was first demonstrated for the similar multi-allelic *bE* and *bW* genes of the *b* mating type locus of *U. maydis* by a series of gene replacements (Gillissen, Bergemann et al. 1992). In *C. cinereus*, this non-allelic interaction was inferred from transformation experiments with genes from the A α locus, where only a single gene-pair exists and one of the specificity genes of the pair may be missing (Casselton 1993).

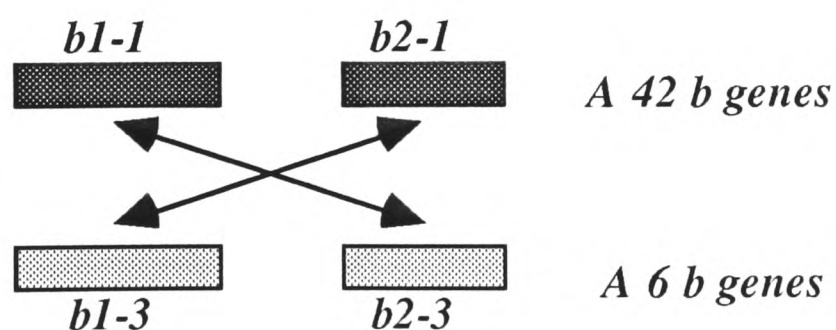


Figure 1: Interaction of the allelic HD1 and HD2 *b* genes of the *C. cinereus* A42 and A6 mating type factors

To prove the rules for this interaction beyond doubt would require gene replacement experiments as possible in *U. maydis* and this has recently been

achieved in *C. cinereus* (Pardo 1993). The mutant strain NA2 was derived from the self compatible mutant A6 strain, in which the only active specificity gene was derived from a fusion of an HD1 and an HD2 gene (see chapter 1). The mutant strain NA2 has had the fusion gene replaced by a disrupted gene. DNA mediated transformation and crosses allowed the introduction of single A mating type specificity genes into this null background in order to study gene to gene interactions. The strategy for introducing the predicted compatible pair of relevance to this study (*b1-3* and *b2-1*) is outlined in figure 6.2. The individual transformants having either *b1-3* or *b2-1* as their only A mating type specificity gene display typical monokaryon morphology and can successfully mate to generate a typical dikaryon.

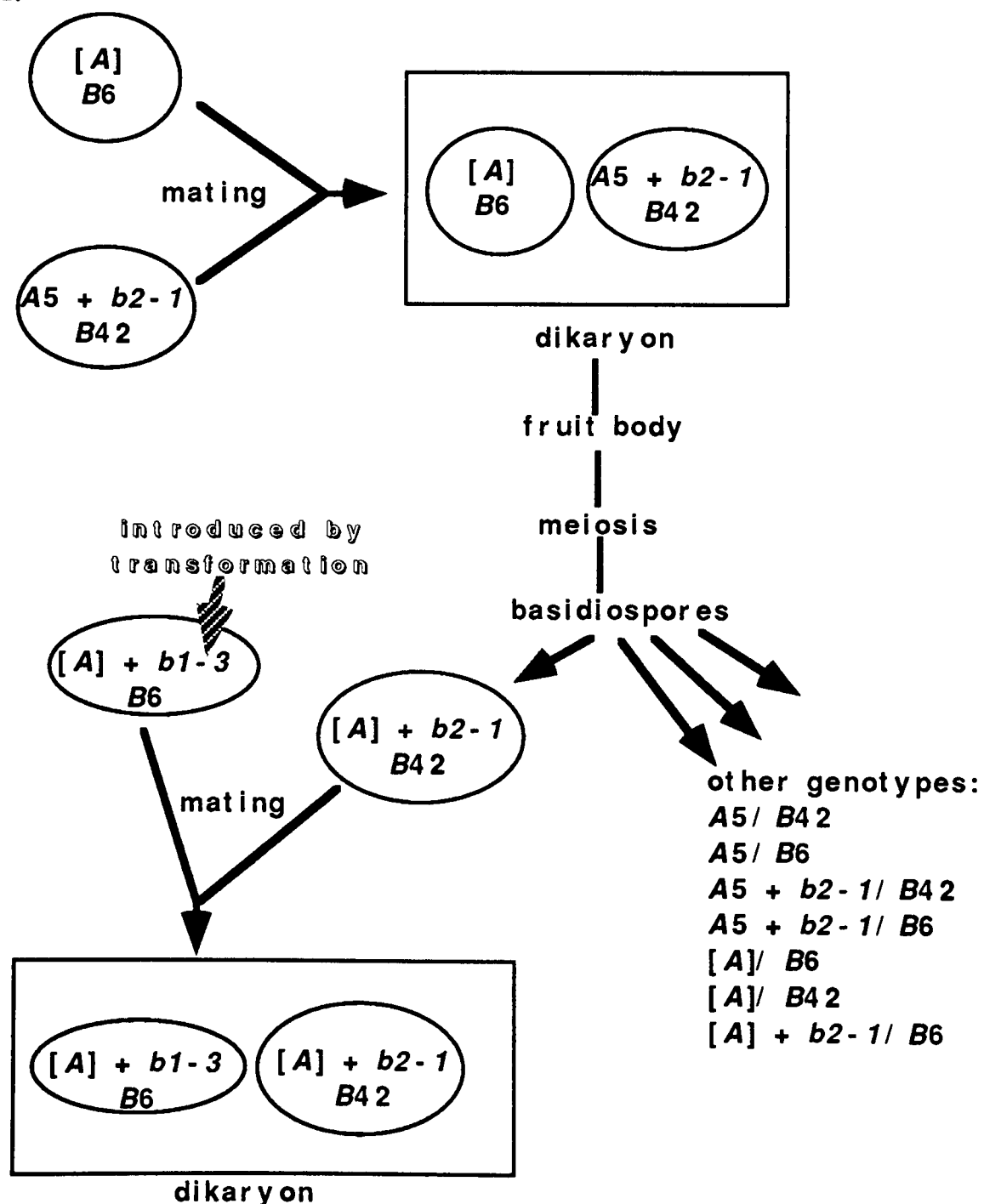


Figure 6.2: Strategy to prove the direct interaction of *b1-3* and *b2-1*. [A] identifies the inactivated A factor of the mutant strain NA2

Having unambiguously established that a compatible A mating type factor interaction is between an HD1 and an HD2 gene coming from different alleles of a HD1/HD2 gene-pair, the experiments described in this chapter are aimed at characterising this interaction on the protein level. The gene products chosen to detect protein-protein interactions were from the *b1-1*, *b2-1* and *b2-3* genes. This would offer the opportunity to compare a compatible reaction (*b1-1* x *b2-3*) and an incompatible reaction (*b1-1* x *b2-1*). The simplest interaction of the HD1 and HD2 proteins would be dimerisation. Although dimerisation has not as yet been shown for the basidiomycete mating type proteins, it has been studied in great detail for the two homeodomain proteins (MAT α_1 and MAT α_2) encoded by the *Saccharomyces cerevisiae* MAT locus (Nasmyth 1982). The α_2 protein is a repressor of a-specific genes in α haploid cells. It binds to the promotor of a-specific genes as an α_2 dimer together with the general transcription factor MCM1 (Nasmyth and Shore 1987). In diploid cells, a novel transcription factor activity is created by formation of heterodimers between the α_1 and the α_2 proteins. This dimer represses the α_1 and other haploid specific genes by binding to their respective promoters.

It has been suggested that the two classes of homeodomain motifs found in the yeast α_1 and α_2 protein are analogous to the two classes of homeodomains found in the basidiomycete mating type proteins. The basidiomycete HD1 type is more closely related to the α_2 homeodomain and the HD2 type more similar to the α_1 homeodomain (Kües and Casselton 1992). By analogy to the yeast system, it was proposed that mating in *C. cinereus* brings together HD1 and HD2 proteins from different alleles of the same gene-pair and it is assumed that an HD1-HD2 dimer is formed following mating. This dimer constitutes a novel transcription factor activity and regulates target genes in order to bring about the phenotypic

switch from typical monokaryotic morphology to a non-sporulating mycelium with unfused clamp cells.

6.1 Analysis of protein-protein interactions by chemical crosslinking

Chemical crosslinking of *in vitro* translated proteins is a standard method for the detection of protein dimerization in solution (Hu, O'Shea et al. 1990). It involves cotranslation of the two relevant RNA species in the presence of ³⁵S-methionine to radioactively label the newly synthesised proteins. This protein mixture is then incubated with a chemical that has two reactive groups separated by a short spacer arm. The crosslinking reagents used in this study were DMSI (di-methyl-suberimidate) and glutaraldehyde which both contain two reactive aldehyde groups that can react with the side chain amino group of lysine to form a Schiff-base adduct. Provided the proteins studied contain lysine side chains at favourable positions, dimerised protein species will be covalently linked. Analysis of the crosslinking reactions by SDS-PAGE should then reveal the appearance of a new protein band having the combined molecular weight of the two proteins contributing to the dimer.

In order to carry out these studies with the b1-1, b2-1 and b2-3 proteins, full length cDNA clones for all three genes had to be available for translation. In the case of the *b1-1* gene, two constructs were available. The full length cDNA in pEb1-1 which has the extended 5'end (see chapter 5) could be used directly as it was constructed in a T7 vector. The cDNA clone pb1-1cDNA isolated originally for the *b1-1* gene (see figure 2.1) could be translated directly as discussed in chapter 5 in order to produce b1-1 protein starting at the 2nd ATG. New full length cDNA constructs had to be prepared for the *b2-1* and the *b2-3* genes.

6.1.1 Construction of a full length *b2-1* cDNA clone

The pb2-1 cDNA clone contains a *b2-1* cDNA insert lacking the first 173 bp of the predicted *b2-1* coding sequence. The missing sequence was added using the strategy shown in figure 6.3. This involved fusing the major part of the cDNA sequence to a fragment of the *b2-1* genomic DNA that lacks introns and extends the gene to 8 bp at the 5' end. This clone was called pb2-1fcDNA.

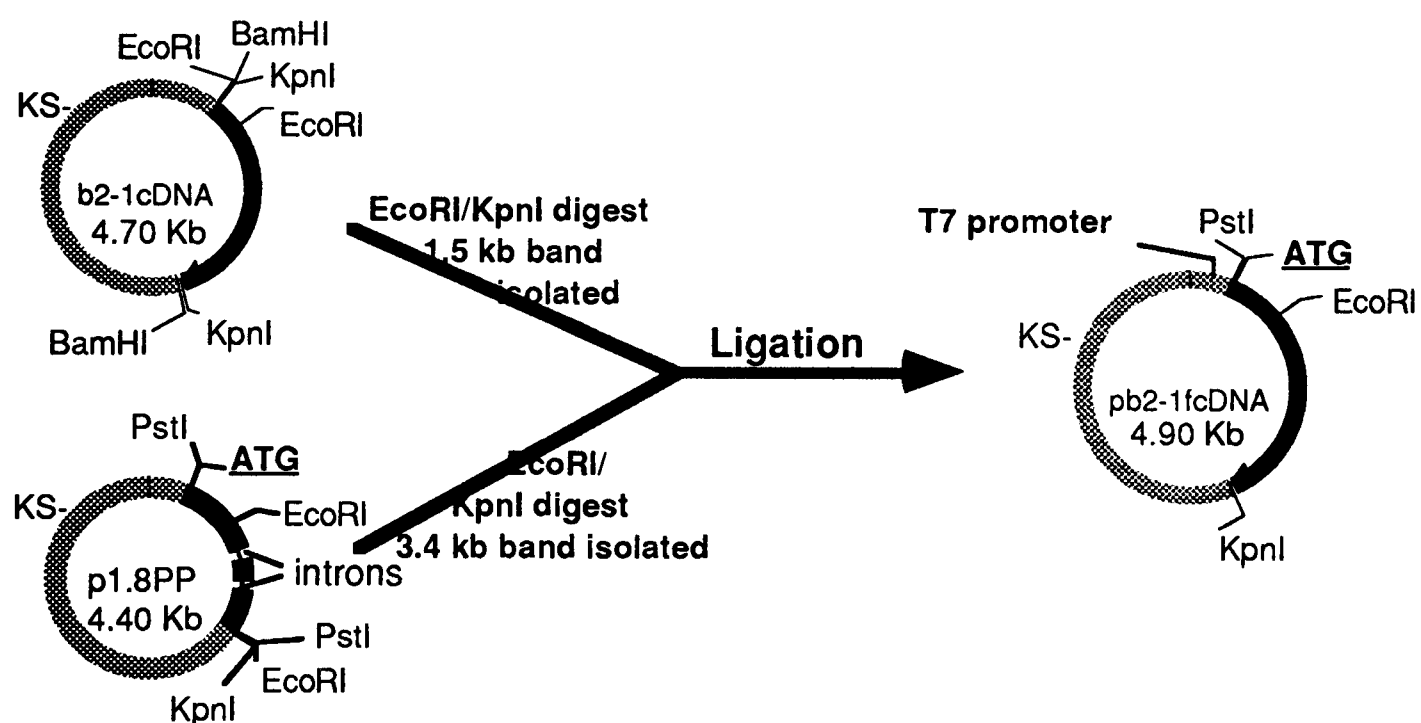


Figure 6.3: Strategy for the construction of pb2-1fcDNA, a full length *b2-1* cDNA clone

Figure 6.4 shows an ethidium stained agarose gel illustrating the cloning strategy used for the construction of pb2-1fcDNA.

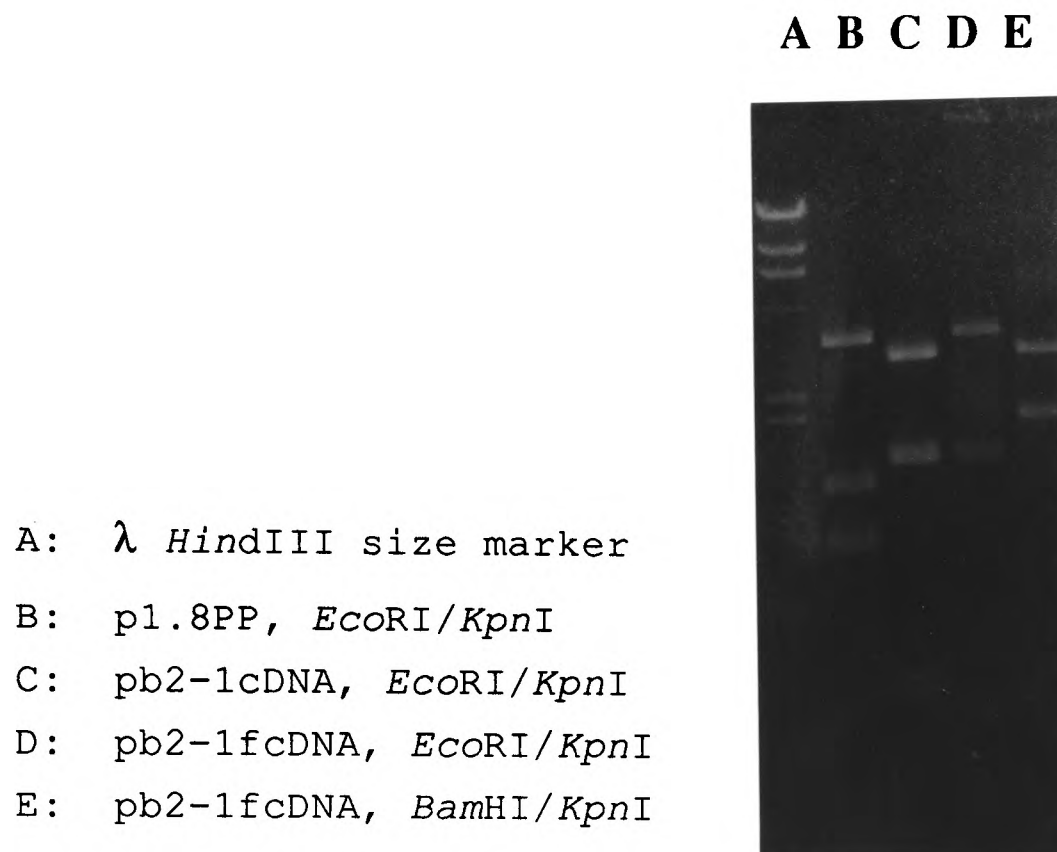


Figure 6.4: Ethidium stained agarose gel with restriction digests of plasmids used for the construction of pb2-1fcDNA.

6.1.2 Construction of a full length *b2-3* cDNA clone for *in vitro* translation

Sequence analysis of the *b2-3* cDNA clones (chapter 4) had shown that pb2-3cDNA2 contained a full length cDNA. This clone could not, however, be used directly for *in vitro* transcription using T7 RNA polymerase because the insert was cloned in the wrong orientation with respect to the T7 promotor. The insert was excised by digestion with *EcoRI/XhoI* and cloned into *EcoRI/XhoI* digested pBluescript KS+. The resulting plasmid, pb2-3cDNA+, contained the *b2-3* cDNA cloned behind the T7 promotor appropriately for *in vitro* transcription. A second suitable full length *b2-3* clone, pT7b2-3, was provided by Dr. A. Banham in this laboratory. This clone had been obtained by engineering a *NdeI* restriction site at the ATG start codon of the *b2-3* gene by PCR and then cloning the resulting PCR product into the T7 expression vector pT7-7. pT7b2-3 was used preferentially for experiments described in this chapter, because it gave higher yields in *in vitro* transcription reactions, presumably due to the fact

that the special T7 expression vector pT7-7 contains all the optimum sequence requirements and spacing for transcription by T7 RNA polymerase.

6.1.3 Chemical crosslinking studies with the b1-1, b2-1 and b2-3 proteins

RNA was translated from the *b2-1*, *b2-3* and the *b1-1* genes using the plasmid constructs just described. The RNA was then *in vitro* translated into ³⁵S-methionine labelled protein. Furthermore, RNAs of the incompatible *b1-1* and *b2-1* and of the compatible *b1-1* and *b2-3* genes were cotranslated to look for heterodimerisation, because it has been shown for the *S. cerevisiae* mating type proteins α_1 and α_2 , that dimerization of *in vitro* translated proteins depends on them being cotranslated (Dranginis 1990). As a positive control, RNA of the *cpf1* gene from *S. cerevisiae* (clone 73-22) was translated and subjected to the same crosslinking procedure because this protein has been shown to homodimerise following *in vitro* translation (Dowell, Tsang et al. 1992). Chemical crosslinking studies were carried out with individually translated proteins to detect possible homodimer formation in addition to the cotranslations designed to detect heterodimers. Two crosslinking agents, DMSI and glutaraldehyde, were used but neither homodimers nor heterodimers were detected, even though the control, which was processed in parallel, produced the expected homodimers. The result of one of these experiments aimed at detecting homodimer formation is shown in figure 6.5, demonstrating that neither *b2-1* nor *b1-1* produce homodimers whereas *cpf1* p73-22 homodimerises as expected.

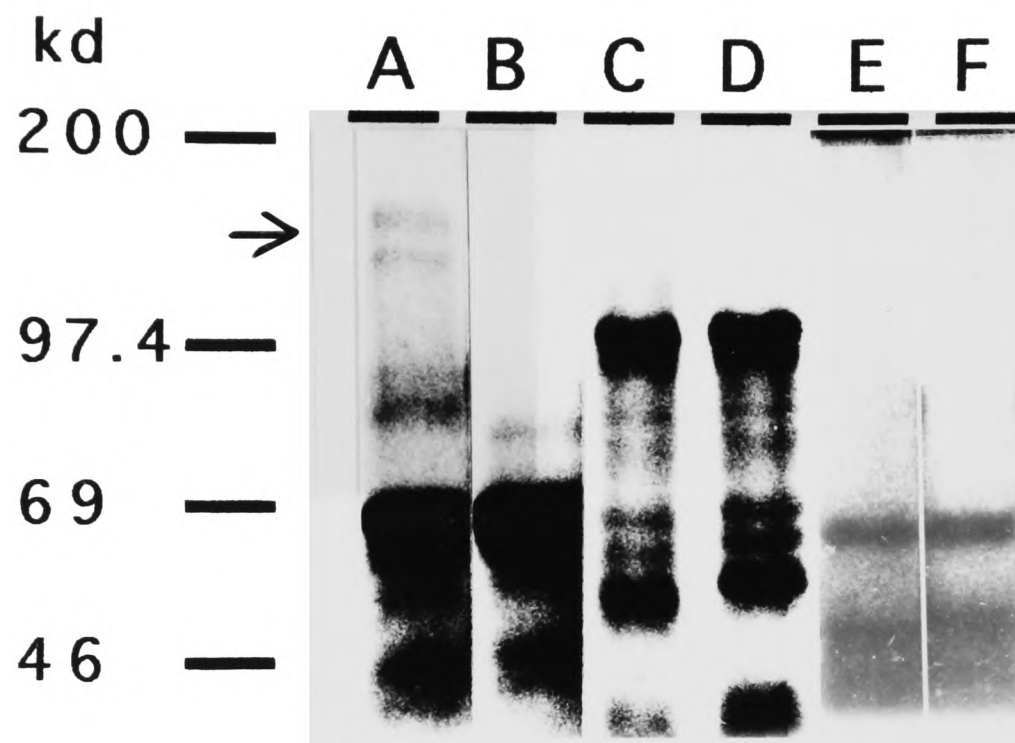


Figure 6.5: Fluorography of SDS-PAGE analysis of a representative chemical crosslinking experiment using DMSI (A= 73-22, B= 73-22 control, C= b1-1, D= b1-1 control, E= b2-1, F= b2-1 control). The cross-linked 73-22 protein can be seen in track (A) as additional bands between the 97.4 and 200 kd size markers (see arrow)

There are two reasons why dimerisation might not have been detected in these experiments. Both crosslinking agents depend on the presence of lysines at favourable positions in the proteins, and these might not have been available in the A proteins tested. For a number of transcription factors, dimerisation of *in vitro* translated proteins depends on the presence of the DNA target site in the assay (Sauer and Jäckle 1993) (Xue, Tu et al. 1993) (Tomei, Cortese et al. 1992) (Ingraham, Flynn et al. 1990). Cooperative binding to DNA may stabilise the dimers or dimer formation is only possible upon DNA binding which might promote a conformational change of the proteins exposing previously hidden dimerisation interfaces.

6.2 Analysis of protein-protein interactions by co-immunoprecipitation

As described in chapter 5, the polyclonal antibodies raised against the A42 b1-1 protein could be used to specifically immunoprecipitate *in vitro* translated b1-1 protein. The second approach used to detect the formation of heterodimers between the b1-1 and the b2-1 or b2-3 protein respectively was by coimmunoprecipitation. The same plasmid constructs were used for cotranslations as in the previous experiment, i.e. pb1-1cDNA for b1-1, pb2-1fcDNA for b2-1 and pT7b2-3 for the b2-3 protein.

Figure 6.6 shows an autoradiograph of an SDS-PAGE analysis of a representative coimmunoprecipitation experiment. The autoradiograph shows that the b1-1 protein is specifically immunoprecipitated by the polyclonal antisera. However, no coimmunoprecipitation of the b2-3 protein could be detected in any of the experiments carried out. In fact, both HD2 proteins (b2-1 and b2-3) were immunoprecipitated by the non-specific pre-immune sera of both rabbits (data shown for b2-3 precipitated by 4923 pre-immune serum in figure 6.6). The highest region of homology between b2-1 and b2-3 is the homeodomain and it might be possible that these immunological results reflect the fact that the *C. cinereus* HD2 proteins possess a typical homeodomain that is found in proteins from other eukaryotes. Both rabbits may already carry antibodies that would react with this region of the HD2 proteins because they may have already been exposed to a very similar epitope.

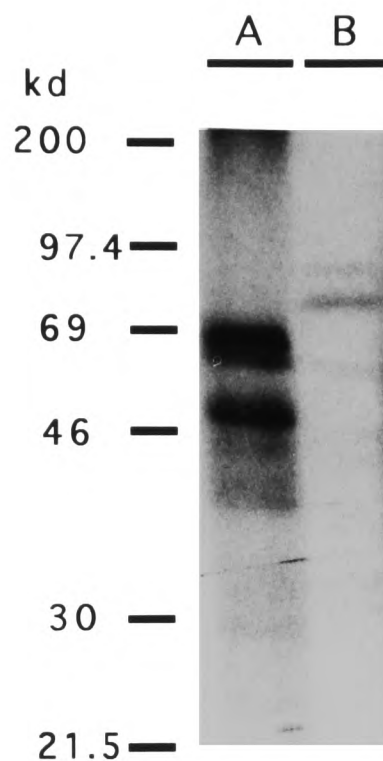


Figure 6.6: Co-immunoprecipitation experiment using co- *in vitro* translated b1-1 and b2-3 protein, analysed by fluorography following SDS-PAGE (track A: pre-immune serum precipitates b2-3; track B: immune serum precipitates b1-1). The bands corresponding to smaller proteins in both tracks A and B were due to premature termination in the *in vitro* reactions as demonstrated by immunoprecipitation experiments using singly translated proteins (data not shown).

Despite this unexpected finding, dimerisation of b1-1 with either the b2-1 or the b2-3 protein under the chosen experimental conditions would have been detected in these coimmunoprecipitation experiments, because the b1-1 protein should be precipitated by the non-specific preimmune sera together with the HD2 protein that it had dimerised with. This was found neither in the b1-1/b2-1 nor in the b1-1/b2-3 coimmunoprecipitation experiments.

Again, failure to detect the predicted dimerisation of the A proteins might have several causes, intrinsic to the experimental procedure rather than the lack of dimerization *in vivo*. Dimerisation may depend on the presence of a DNA target site in the assay as shown for a number of transcription

factors (Sauer and Jäckle 1993) (Xue, Tu et al. 1993) (Tomei, Cortese et al. 1992) (Ingraham, Flynn et al. 1990). Alternatively, the proteins might not be folded correctly after *in vitro* translation in order to allow dimerisation . Dimerisation *in vivo* might even only be possible after some post-translational modifications of either of the proteins or may require a third auxiliary factor. Protein-protein interactions depend on the ionic composition of the solution that they are assayed in. The high protein concentrations present in the cells *in vivo* cannot be mimicked in the *in vitro* crosslinking or coimmunoprecipitation experiments. Both these experiments demand dilution of the assay in order to increase specificity and reduce, for instance, the accidental crosslinking of two proteins caused by "molecular crowding".

6.3 Detection of protein-protein interactions using b1-1-GST fusion proteins and *in vitro* translated *C. cinereus* A mating type proteins

Successful demonstration of protein-protein interactions between the *C. cinereus* A mating type proteins was achieved by a new technique that has been shown to be very sensitive at detecting protein-protein interactions between transcription factors (Yuan, Stroke et al. 1993) (Sessa, Morelli et al. 1993) (Andres, Chiara et al. 1994) (Hsu, Cheng et al. 1991). The technique, described only recently, has been called the GST-fusion protein *in vitro* association assay (Kerr, Ransone et al. 1993). It involves the association of a bacterially produced GST fusion protein with an interaction partner that has been produced by *in vitro* translation. The GST fusion protein is immobilised by binding to a glutathione sepharose resin via its GST component and then incubated with the *in vitro* translated interaction partner.

Interactions between the two proteins lead to an indirect binding of the *in vitro* translated protein to the resin via its association with the GST fusion protein. The supernatant fraction is removed, the resin fraction is washed a few times and finally the supernatant and the resin fraction are analysed by SDS-PAGE. The gels are dried and exposed to X-ray film. If the *in vitro* translated protein which is labelled with ^{35}S -methionine can be detected in the resin fraction, then a specific protein-protein interaction between the *in vitro* translated protein and the GST fusion protein has occurred. A parallel experiment is carried out in which the GST protein (only the GST part of any fusion protein) is bound to the glutathione sepharose beads and incubated with an *in vitro* translated protein. This would detect any non-specific interactions either with the resin or with the GST part of a GST fusion protein. The experimental approach is outlined in figure 6.7.

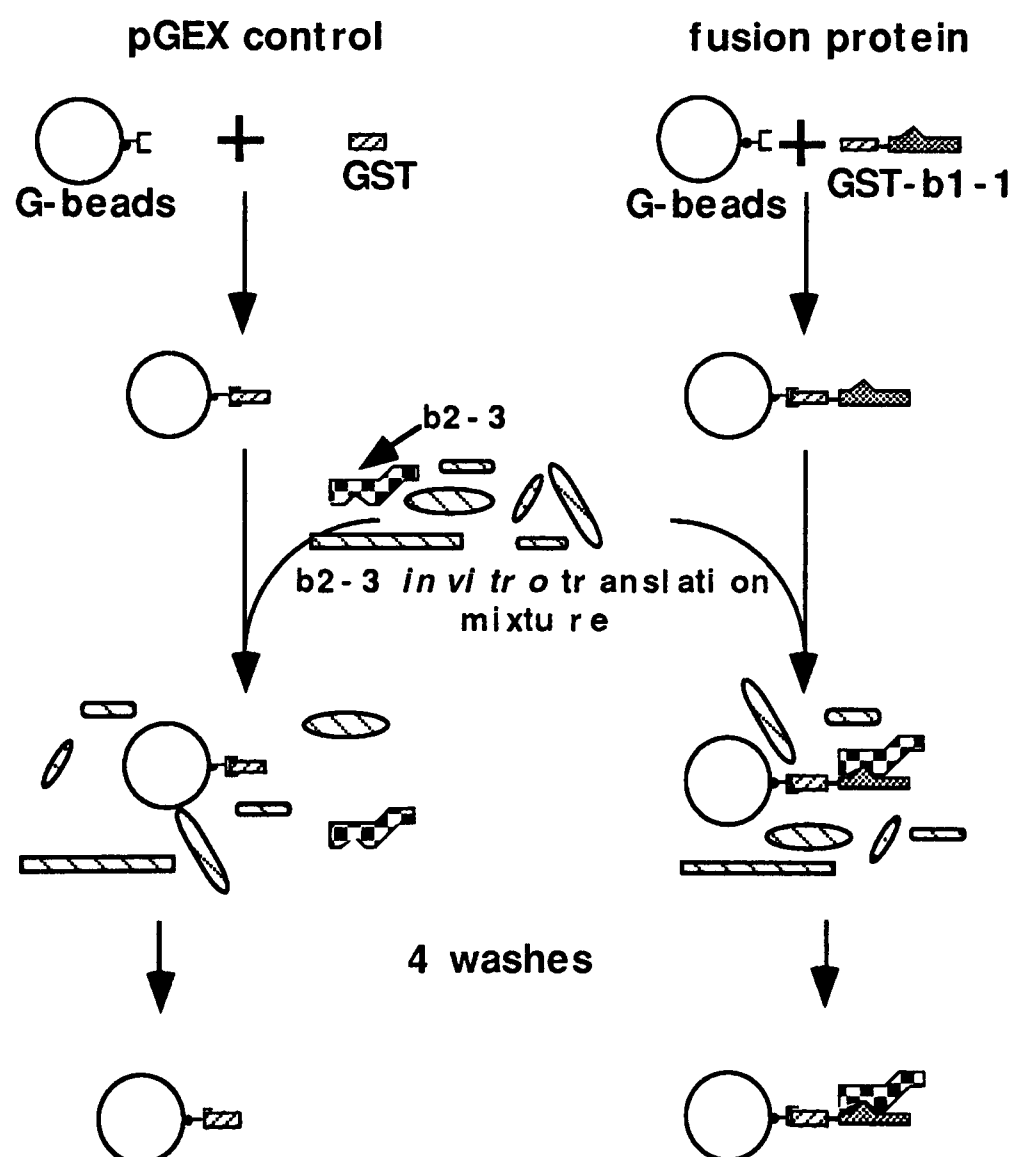


Figure 6.7: Principle of glutathione fusion protein association experiment

6.3.1 Initial XB1 b2-3 association experiment

In the first GST association experiment, the XB1 fusion protein containing the N-terminal region of b1-1 and *in vitro* translated b2-3 protein were tested for a specific interaction by incubating both proteins in 1x PBS and washing the resin fraction four times with 1x PBS. Figure 6.8 shows, that under these conditions the b2-3 protein was found to bind non-specifically to the resin fraction with or without the XB1 protein.

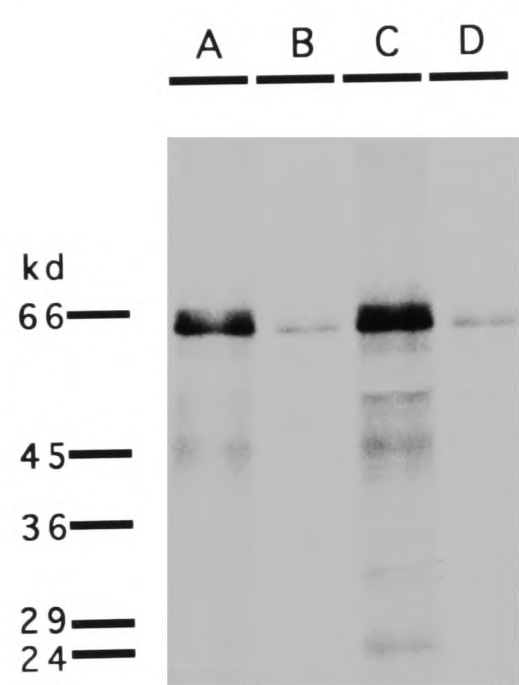


Figure 6.8: Initial XB1/b2-3 association experiment. XB1 and b2-3 were incubated in 1 x PBS and the resin fraction was washed four times with PBS. Fluorography of SDS-PAGE shows that b2-3 is found in the resin fractions for the XB1 and the GST control experiment (A= GST control resin, B= GST control supernatant, C= XB1 resin, D= XB1 supernatant).

The specificity of the assay therefore needed to be increased in order to distinguish a specific association between the XB1 (b1-1) and b2-3 proteins from non-specific interactions between GST (or even just the glutathione resin) and b2-3. The assay was modified in two ways. Firstly, BSA at 10 mg/ml was added throughout the incubation in order to block any non-specific affinity for proteins by either the GST or the glutathione sepharose matrix. Secondly, DTT was added to the incubation solution at a final concentration of 7.5 mM in order to keep all the SH groups of the

glutathione sepharose and the GST protein in their reduced SH state thereby preventing the b2-3 protein from binding by means of disulphide bridges. The method is described in detail in chapter 2.

6.3.2 Identification of potential dimerisation regions of the b1-1 and the b2-3 proteins using the GST association assay

Using the more stringent assay conditions, a clear protein-protein interaction between the b1-1 and the b2-3 proteins was demonstrated using the GST association assay with the XB1 protein immobilised on the glutathione sepharose matrix, incubated with *in vitro* translated b2-3 protein. The result is shown in figure 6.9. It is apparent, that there is very little residual non-specific association of the b2-3 protein to the resin fraction in the GST control compared with the strong binding seen when b1-1 sequences are present. The modifications to the method, i.e. addition of BSA and DTT to the assay, have thus almost completely inhibited the non-specific binding.

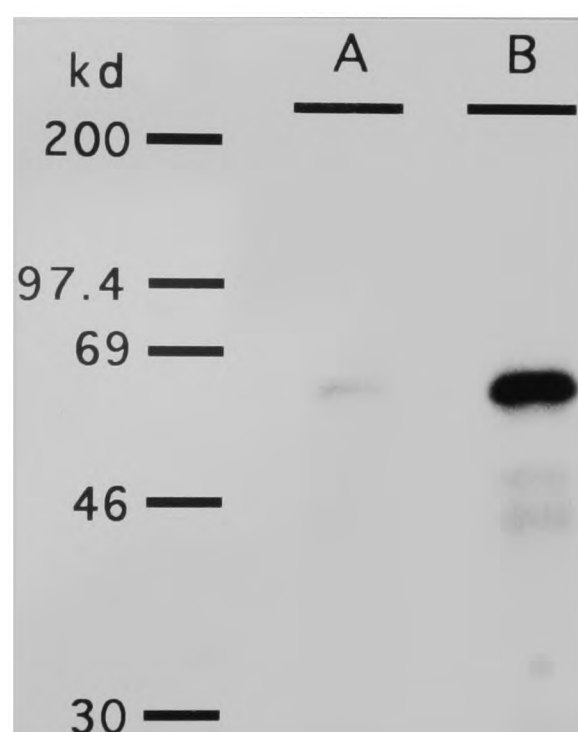


Figure 6.9: XB1/b2-3 association experiment analysed by fluorography following SDS-PAGE (A= GST control resin fraction, B= XB1 resin fraction).

The experiment described above showed a clear specific protein-protein interaction between the b1-1 and the b2-3 proteins which was not possible to demonstrate by chemical cross-linking and co-immunoprecipitation experiments using *in vitro* translated proteins.

The higher sensitivity of the GST association experiment may be explained by the fact that one can use an excessive amount of the GST fusion protein to sequester the much smaller amount of *in vitro* translated association partner. With both proteins being translated *in vitro*, successful associations may simply be below the detection level of the assay used. Other factors may be: 1.) The addition of 10 mg/ml BSA approaches the high protein concentrations to be expected in the nucleus, and this may in some way promote specific interactions of the *A* mating type proteins. 2.) The GST fusion protein is immobilised on the glutathione sepharose matrix thereby reducing the energy barrier necessary for binding between the two proteins. The free energy at a protein interface required for dimerization in solution must be high enough to overcome the loss of translational and rotational degrees of freedom of the two proteins (Harrison and Aggarwal 1990). With one protein being immobilised, this energy barrier should be significantly lower. As mentioned in section 6.1, dimerisation between transcription factors could often only be detected if the DNA target sites are included in the assay. In some respect, immobilising one of the interacting proteins on the glutathione sepharose matrix may partly mimic the effect of the DNA target site by immobilising the protein and at least partly locking it into a particular folding structure. 3.) It has been shown that the *S. cerevisiae* mating type proteins α_1 and α_2 dimerise more readily *in vitro* if the proteins were produced in *E. coli* rather than by *in vitro* translation (Goutte and Johnson 1993). The *C. cinereus* mating type proteins may have analogous properties, and the GST association experiment uses one interaction partner that was produced in *E. coli*.

The experiment was repeated using the shorter XC1 fusion protein and *in vitro* translated b2-3. The XC1 protein contains the region of the b1-1 protein N-terminal to the third helix of the homeodomain, 50 amino acids shorter than XB1. The results were identical for both experiments and it can be concluded that the first 162 amino acids of the b1-1 protein are sufficient to mediate a specific protein-protein interaction with the b2-3 protein.

To define the essential interacting region of the b2-3 protein more precisely, the b2-3 clone (pT7b2-3) was linearised with various restriction enzymes before *in vitro* transcription. This produced C-terminally truncated versions of the b2-3 protein following *in vitro* translation. In addition, a new expression construct was prepared using the pb2-3cDNA1 clone. This cDNA clone lacks the first 72 nucleotides of the b2-3 coding sequence which removes the N-terminal 24 amino acids of the b2-3 protein. The cDNA was cloned into the T7 expression vector pT7-7 to provide an ATG start codon and the T7 promoter required for *in vitro* transcription/translation. The resulting clone pT7b2-3a was digested with *Nco*I and used for *in vitro* transcription/translation. The translation product (b2-3 Δ NNco) showed the same specific protein-protein interaction with the XC1 fusion protein as the full length b2-3 protein (see figure 6.10).

The results of these experiments are summarised in figure 6.10. It is apparent that the majority of the C-terminus of the b2-3 protein can be deleted without loss of the interaction with the XC1 protein. Similarly, the N-terminal 24 amino acids are not required for this protein-protein interaction either, thus narrowing down the interacting region to amino acids 25 to 115.

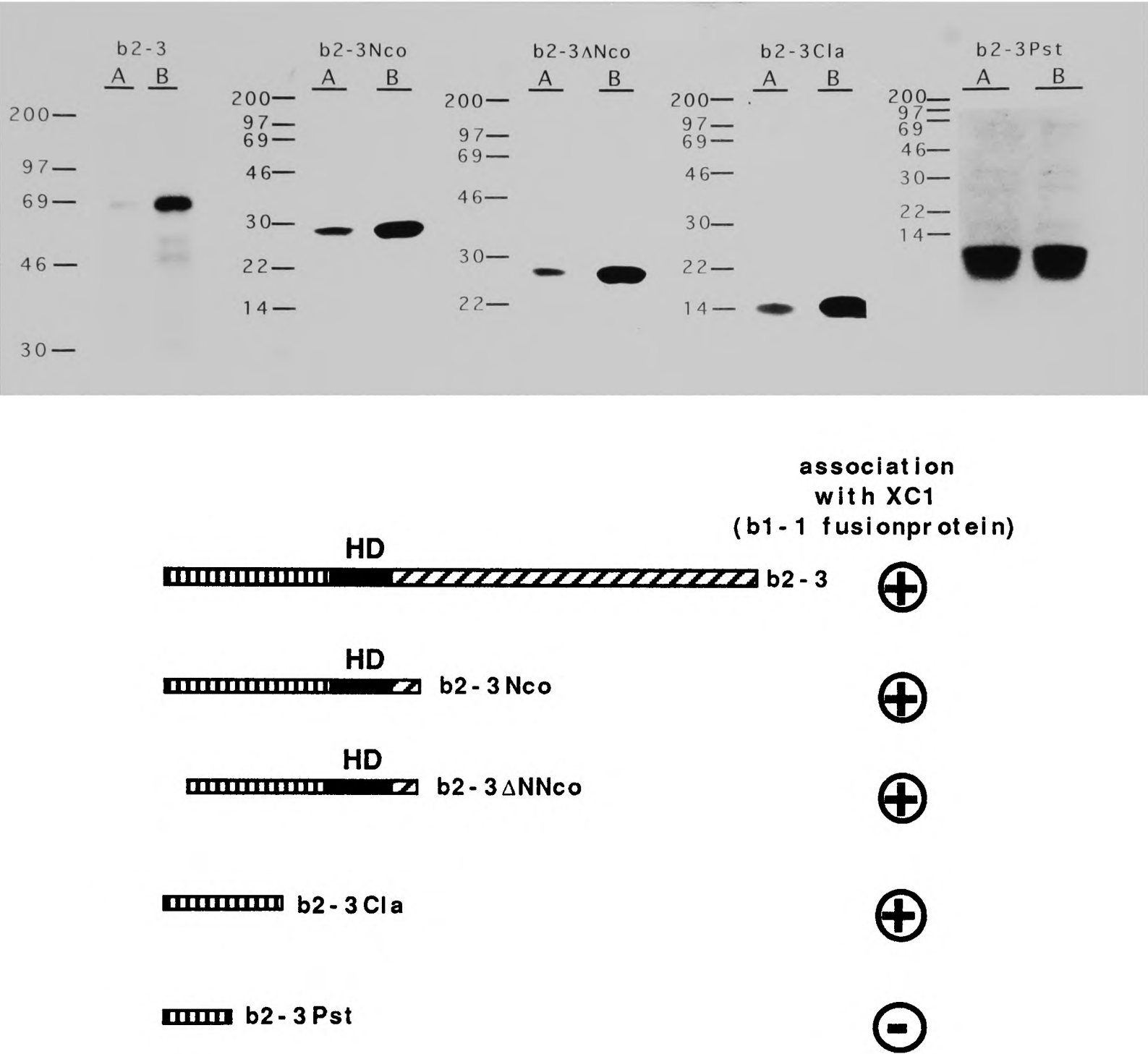


Figure 6.10: XC1 association assay with C-terminal truncations of the b2-3 protein. Shown are the resin fractions of the GST control and the XC1 fusion protein experiments for all truncations and a sketch illustrating the position of truncated proteins with respect to the full length b2-3 protein (the last panel = b2-3Pst shows band intensities after a prolonged exposure).

The smallest fragment of the b2-3 protein that still showed the interaction with the XC1 fusion protein was b2-3 Cla (see figure 6.10) which contains

the N-terminal 115 amino acids of b2-3. The b2-3 Pst fragment containing the first 73 amino acids, did not show the interaction. This implies that the region between amino acids 73 and 115 is essential for the interaction between b1-1 and b2-3. The amino acid sequence of this region will be examined in more detail in chapter 8 where it will be compared with the corresponding region of the allelic b2-1 protein.

6.3.3 Assay for interactions between the XB1 and *in vitro* translated b1-1 protein using the GST association technique

The GST *in vitro* association assay was used to look for interactions between *in vitro* translated b1-1 protein and the XB1 fusion protein. This interaction would constitute homodimerisation between the b1-1 proteins. The homeodomain of the b1-1 protein has been classified as a basidiomycete HD1 type (Kües, Richardson et al. 1992) and it has been suggested that it has limited analogy to the *S. cerevisiae* MAT α 2 homeodomain (Kües and Casselton 1992). As discussed in the introduction, the MAT α 2 protein represses a-specific genes by binding to a specific upstream regulatory site as a homodimer in conjunction with the MCM1 protein (Smith and Johnson 1992). The experiment described here questions whether analogous homodimers can be formed by the *C. cinereus* A mating type HD1 proteins.

The clone used for *in vitro* translation of *b1-1* was pb1-1cDNA (see figure 2.1). The association experiment was carried out using the XB1 fusion protein which contains the N-terminal 212 amino acids of the b1-1 protein including the homeodomain. The result is shown in figure 6.11. No interaction between the XB1 and the b1-1 proteins could be detected and in both experimental and control panels it can be seen that the b1-1 protein is found exclusively in the supernatant fraction.

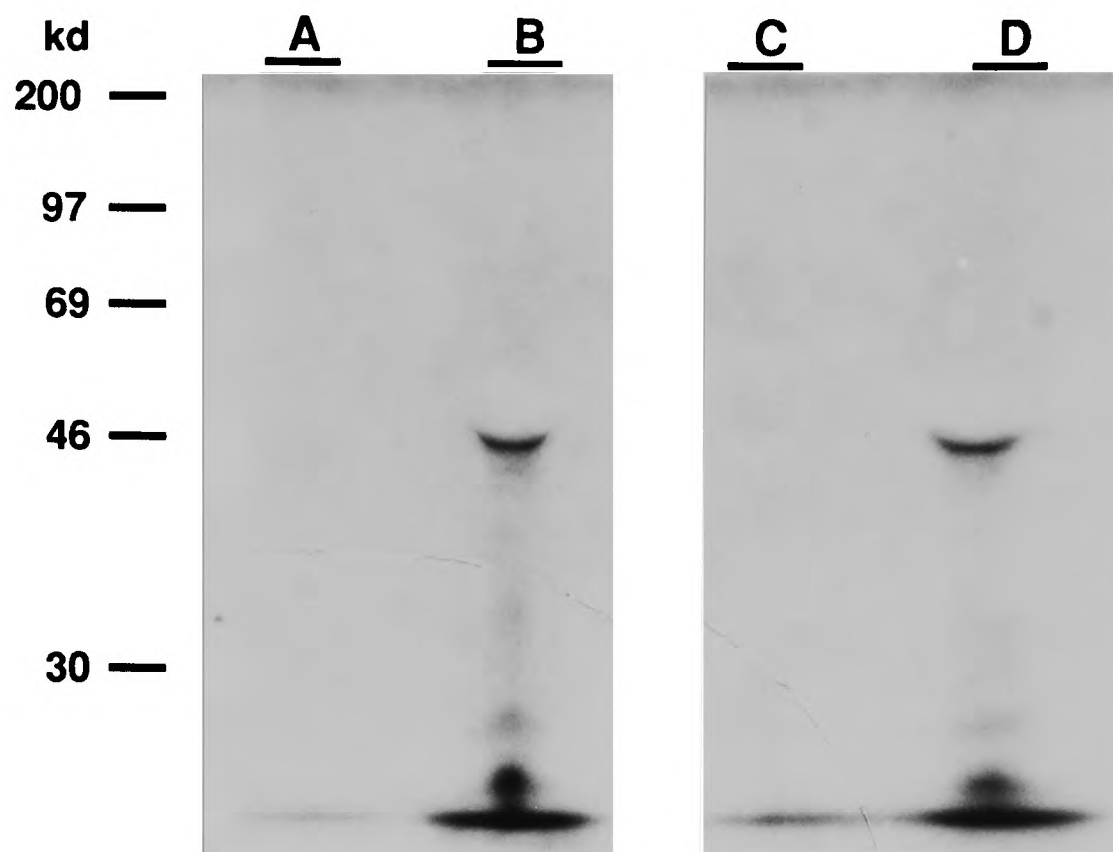


Figure 6.11: GST association assay using the XB1 fusion protein and *in vitro* translated b1-1. The b1-1 protein does not appear at its expected size (approximately 90 kD) in the supernatant fractions because these tracks are heavily overloaded with BSA (170 μ g per track) which interferes with the proper size separation in the higher molecular weight range (A= GST control resin, B= GST control supernatant, C= XB1 experiment resin, D= XB1 experiment supernatant).

The result of the XB1/b1-1 association experiment indicates that there is no homodimerisation of the HD1 proteins of *C. cinereus*, at least not via their N-terminal region. It must be remembered however, that the XB1 protein used in this assay is a truncated version of the b1-1 protein and thus the experiment could not explore possible dimerisation interfaces in the C-terminal half of b1-1.

6.3.4 Analysis of XB1/b2-1 protein-protein interactions using the GST *in vitro* association assay

An important property of the A mating type HD1 and HD2 proteins is the ability to discriminate between different allelic forms which can interact to promote sexual development. To examine whether the association between b1-1 and b2-3 demonstrated in section 6.3 is allele specific, the GST association assay was carried out using the XB1 protein and *in vitro* translated b2-1 derived from the clone pb2-1fcDNA (see section 6.1.1). Figure 6.12 shows that no specific interaction between the b2-1 and the XB1 fusion protein could be detected.

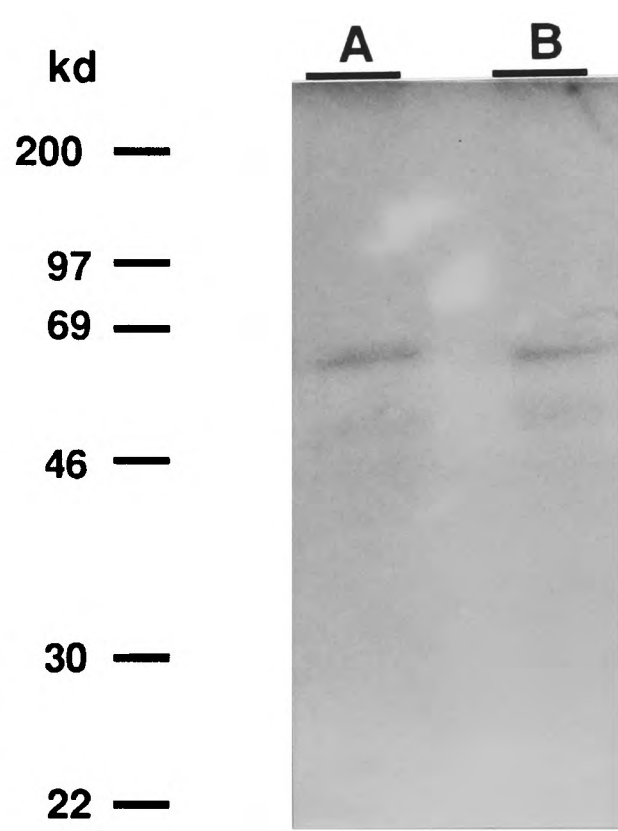


Figure 6.12: XB1/b2-1 association experiment. Fluorography of SDS-PAGE analysis (A= GST control resin fraction, B= XB1 experiment resin fraction).

Experiments using chimeric genes identified the N-terminal regions of both HD1 and HD2 proteins as being the determinants of a compatible HD1-HD2 interaction (see chapter 4). The experiments described in this chapter suggest a molecular mechanism for this allele specific interaction:

compatible interactions are recognised by dimerization, incompatible HD1 and HD2 proteins cannot dimerise. The experiments show that b1-1 is unable to interact with b2-1 but can interact with b2-3 and, one must assume, all other compatible b2 proteins. In unmated monokaryons, only incompatible proteins are present. Mating brings together different alleles of the genes which when translated result in the dimerisation between compatible HD1 and HD2 proteins. This dimerisation generates a new active transcription factor activity that promotes the switch from monokaryotic to dikaryotic morphology.

Chapter 7

Expression of the *C. cinereus* A mating type proteins in *S. cerevisiae*

The experiments described in the previous chapter demonstrated that the N-terminus of the b1-1 protein can interact with the N-terminus of the b2-3 protein. Both proteins contain a homeodomain motif downstream of this putative dimerisation region which is thought to mediate the DNA binding of the A proteins. The C-terminal half of the b1-1 protein has been shown to function as a strong transactivation domain in *S. cerevisiae* (Banham 1993). These data made it possible to predict the function of b1-1 and b2-3 in a compatible mating reaction. After hyphal fusion of an A42 and an A6 monokaryon, b1-1 and b2-3 dimerise. The novel heterodimer binds to the operator region of genes that are responsible for A regulated sexual development. The transactivation domain of b1-1 only activates transcription of A target genes when bound as a heterodimer. One could speculate that b1-1 with its atypical homeodomain can only bind to the operator as a heterodimer with a second protein with a typical DNA binding domain, i.e. the HD2 of b2-3. The experiments described in this chapter were designed to establish an experimental system to study protein protein interactions of b1-1 and b2-3 at potential DNA target sites. The experiments were carried out using the heterologous host *S. cerevisiae*.

S. cerevisiae is well studied at a molecular level and has been used as a model system to analyse eukaryotic protein function. Numerous mutants and vector systems exist to clone foreign genes and study their function. It has been shown that proteins expressed in *S. cerevisiae* may retain their function and cloned eukaryotic genes have been identified by complementing mutations in *S. cerevisiae* (Pillai, Kythe et al. 1993). The

two hybrid system (Chien, Bartel et al. 1991) allows the analysis of protein-protein interactions *in vivo* in the heterologous yeast system (Fritz and Green 1992). The enhancer trapping technique (Chang and Timberlake 1992) has shown that DNA binding proteins retain their ability to bind DNA at high sequence specificity when expressed in *S. cerevisiae*. This approach could therefore be used to identify DNA target sites. The properties of the *b1-1* and *b2-3* proteins described above meet all the requirements for experiments following the enhancer trapping approach, i.e. protein-protein interactions resulting in sequence specific DNA binding of a heterodimeric protein complex that contains a strong transactivation domain in yeast.

There are many homeodomain proteins and other transcription factors described in the literature that are thought to control their own rate of expression (Bienz 1994) (Kanaan and Marzluf 1993) (Chiao, Miyamoto et al. 1994) and that have been found to bind to their own operators (Beachy, Varkey et al. 1993) (Zappavigna, Renucci et al. 1991). It was reasonable to analyse whether a similar regulatory control is operating for the *C. cinereus* *A* mating type proteins. The system chosen for this analysis was a multi vector reporter gene system in *S. cerevisiae* which has been used for the enhancer trap approach to identify DNA target sites of DNA binding proteins (Chang and Timberlake 1992). The experimental approach is illustrated in figure 7.1.

The *b1-1* and *b2-3* genes were cloned into the yeast expression vectors pAA55 and pAA54 respectively. Expression of both genes depended on growth in medium containing galactose as major carbon source. The intragenic region between the *b1-3* and *b2-3* genes was cloned into the reporter plasmids pYC7 and pYC8. These vectors contain a multiple cloning site upstream of a minimal *cyc1* promotor and a *lacZ* reporter gene. If all

plasmids have been successfully introduced into the same cell and if the intragenic region of the A6 *b* genes contains a binding site for the b1-1/b2-3 proteins, then expression of these genes induced by growth on galactose would result in their binding to the sequence cloned in pYC7/pYC8. The transactivation domain of b1-1 is known to be able to activate transcription of the reporter gene and this would result in blue colour development on media containing the chromogenic substrate X-Gal. The system was chosen because it could be used in future to screen for potential DNA binding sites for the A mating type genes by cloning random genomic DNA fragments into the pYC8 reporter vector.

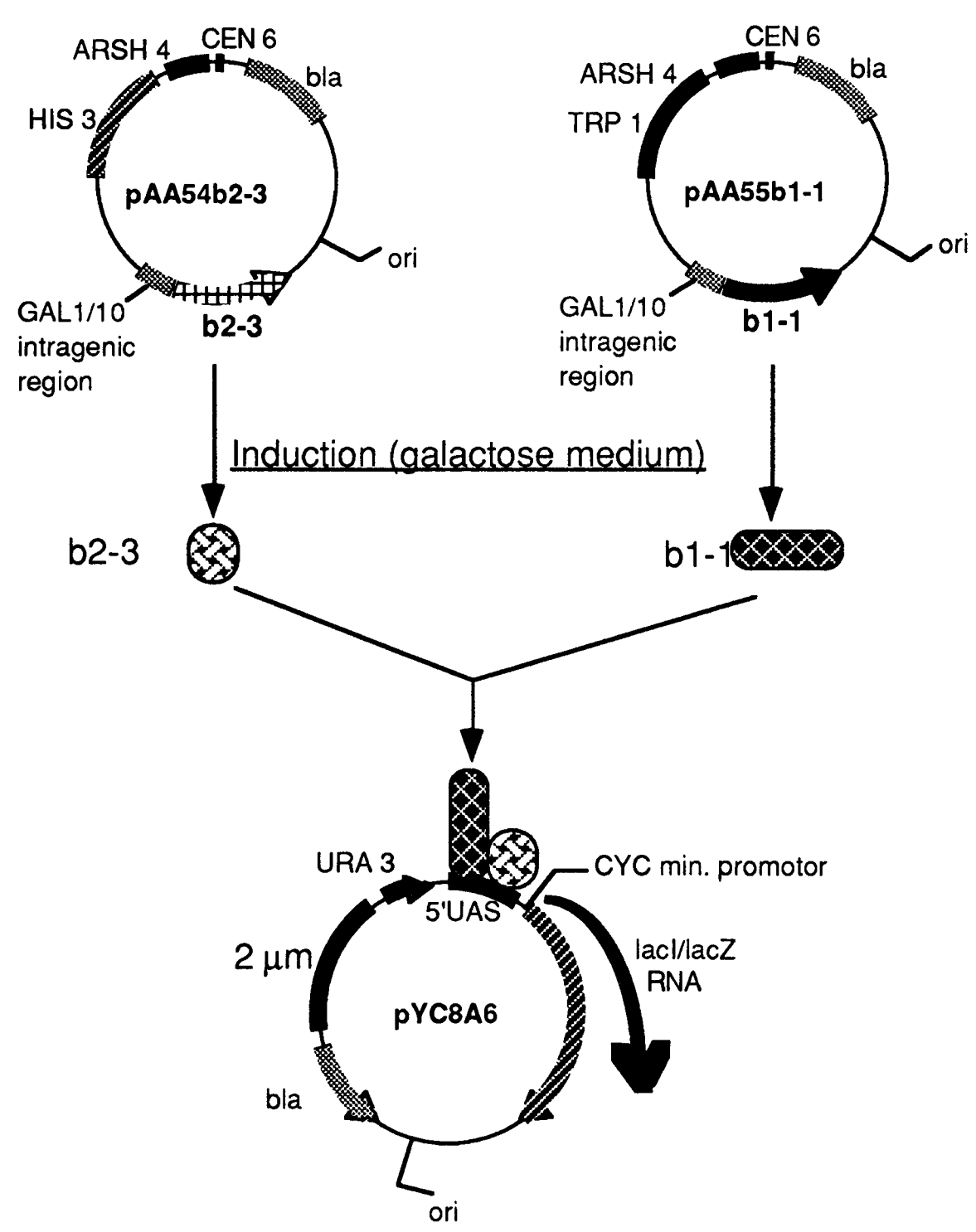


Figure 7.1: Principle of the enhancer trapping technique using a three vector system in *S. cerevisiae*

7.1 Construction of plasmids

A full length *b2-3*cDNA clone (pb2-3AB) prepared *in vitro* by Dr. A. Banham was linearised with *Bam*HI and religated after filling in the overhangs with Klenow. After transformation, all clones analysed had lost the *Bam*HI site as expected. The resulting clones were called pb2-3fs and digested with *Spe*I, overhangs were filled in with Klenow enzyme and DNA was digested with *Xho*I and the resulting 1.8 kbp band was gel-purified.

pAA54 DNA was digested with *Nco*I and the overhangs were filled in. The DNA was then digested with *Sal*I and gel-purified. The *b2-3* and the pAA54 fragments were ligated to yield pAA54b2-3 (cloning strategy shown in figure 7.2). The resulting in-frame fusion of the *b2-3* coding sequence behind the GAL1 start codon is illustrated in figure 7.3.

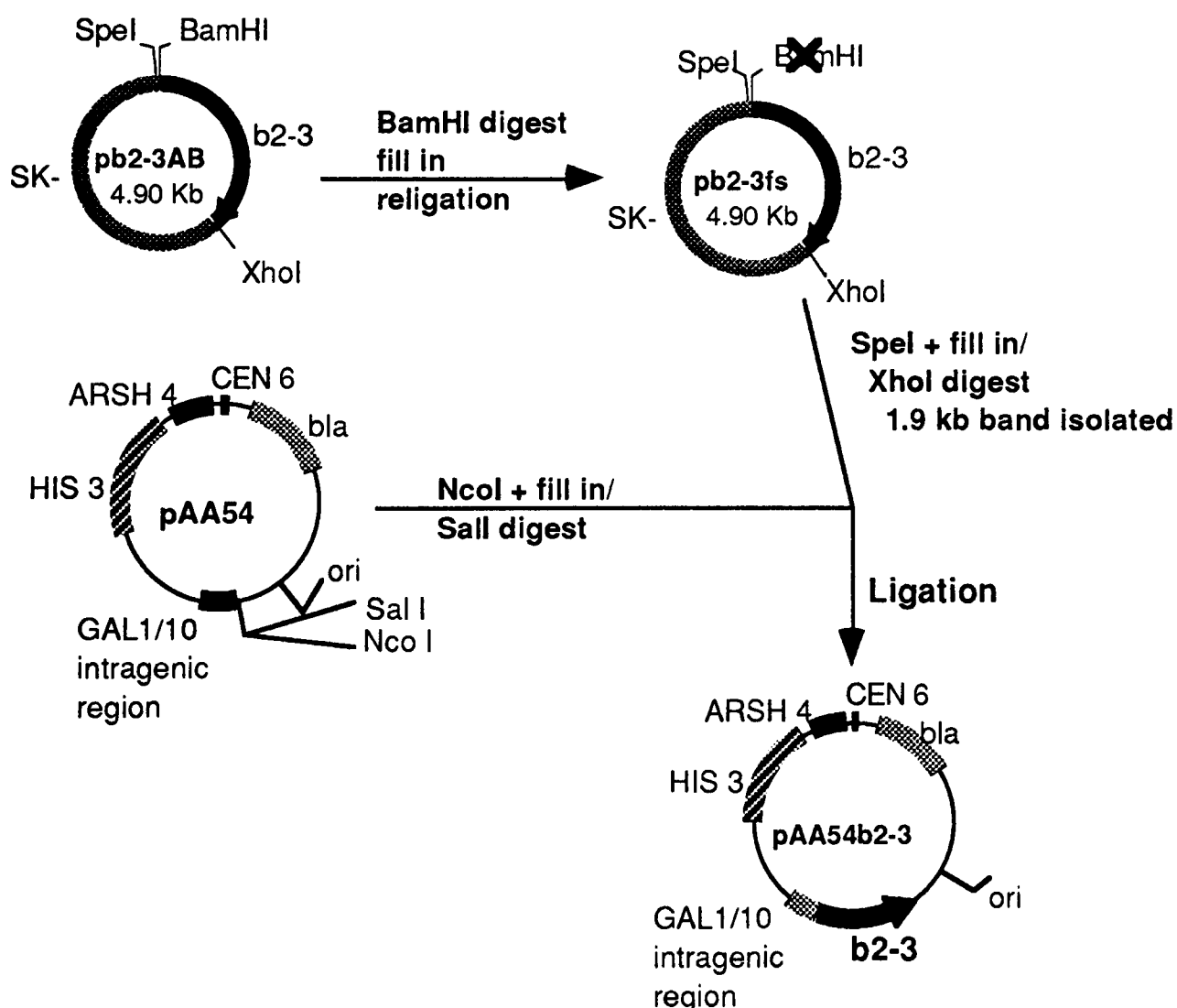


Figure 7.2: Cloning strategy for the construction of the *b2-3* expression construct pAA54b2-3. **X** indicates restriction sites lost due to fill-in reactions.

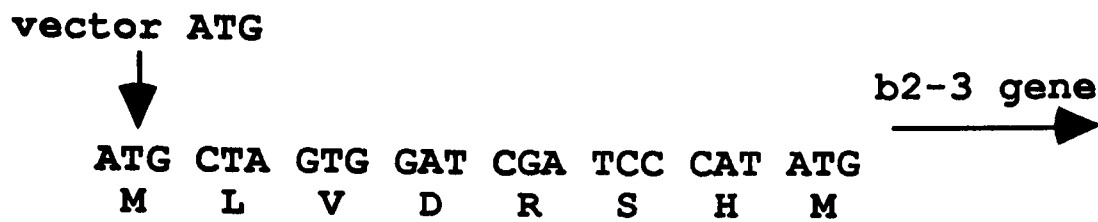


Figure 7.3: In frame fusion of the *b2-3* coding sequence in the pAA54b2-3 expression construct

The expression construct for *b1-1* was prepared in a similar way. *pb1-1*cDNA (figure 2.1) was digested with *Sma*I/*Xho*I and the resulting 2 kbp fragment was gel-purified. pAA55 DNA was digested with *Nco*I and overhangs were filled in. The DNA was then digested with *Sal*I, gel purified and ligated to the *b1-1*cDNA fragment to give pAA55b1-1. The cloning strategy is illustrated in figure 7.4. The in-frame fusion of the *b1-1* start codon to the *GAL1* start codon of the pAA55 expression vector is shown in figure 7.5.

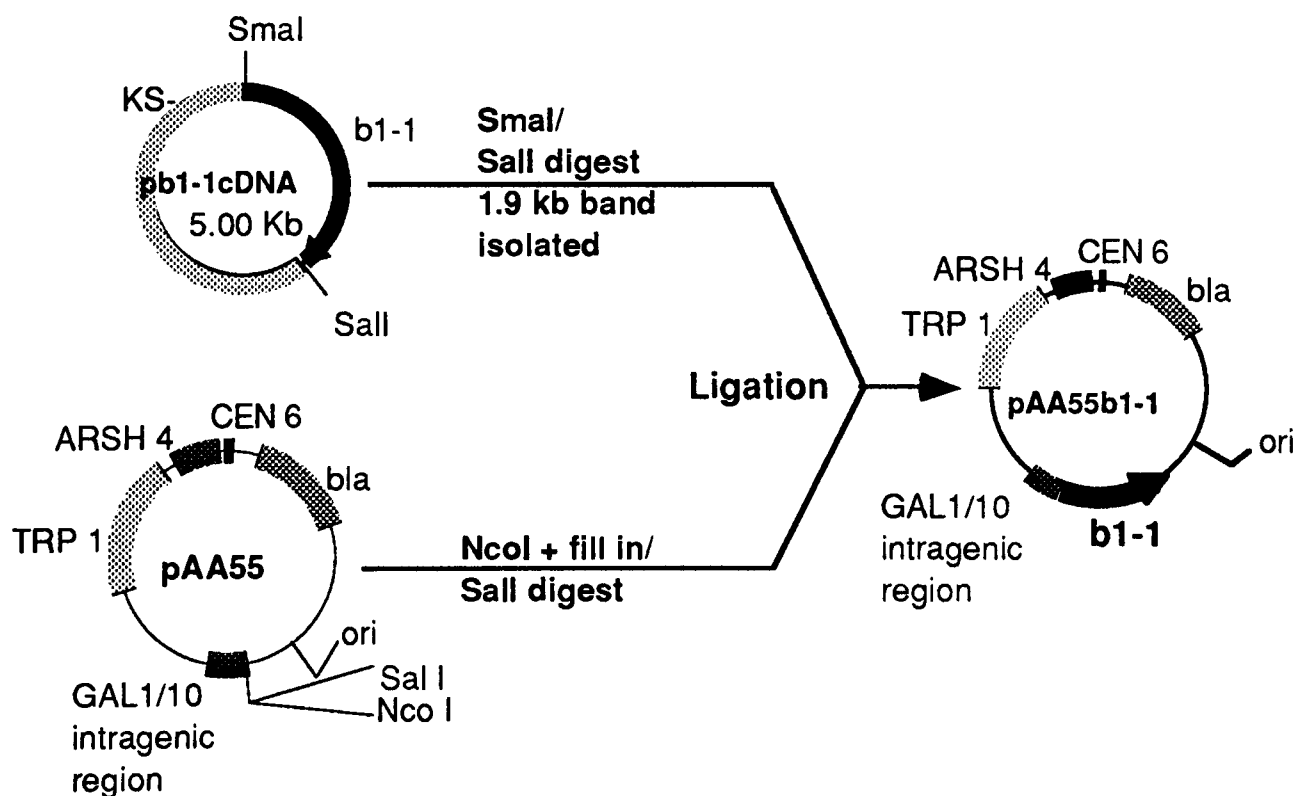


Figure 7.4: Cloning strategy for the construction of the pAA55b1-1

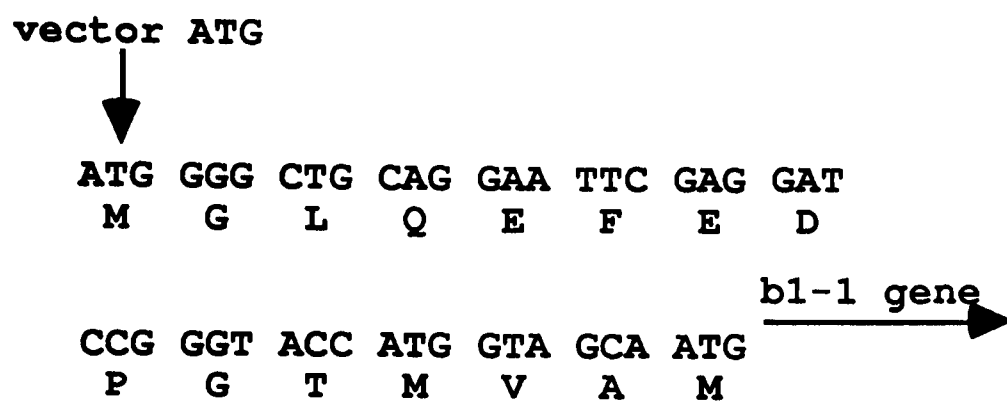


Figure 7.5: In frame fusion of the *b1-1* coding sequence in the pAA55b1-1 expression construct

The intragenic region of the A6 *b* gene-pair was cloned into the vectors pYC7 and pYC8 which allows cloning of putative regulatory sequences upstream of the minimal CYC1 promotor-*lac* Z fusion in either orientations. Figure 7.7 shows a map of the vector pYC8. The vector pYC7 is identical but with the multiple cloning site in the opposite orientation. pA6211 DNA was digested with *Eco*RI and overhangs were filled in. After digestion with *Sal*I, a 0.9 kbp *Sal*/blunt fragment was gel-purified (from position 2556 to 3494 in the A6 *b* gene-pair sequence, see appendix). Figure 7.6 shows the position of this fragment with respect to the start of the *b1-3* and *b2-3* genes. This fragment was cloned into *Sma*I/*Sal*I digested pYC7 and pYC8 to yield the clones pYC7A6 and pYC8A6.

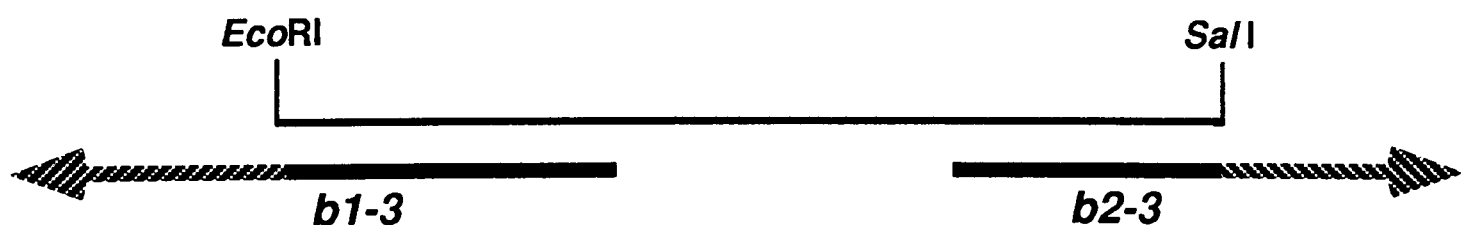


Figure 7.6: Location of intragenic region of the A6 *b* gene-pair cloned into pYC7/pYC8 with respect to the position of *b1-3* and *b2-3*. The shaded portions of the *b1-3* and *b2-3* genes are not drawn to scale.

Figure 7.8 shows an ethidium stained agarose gel with digests of various clones illustrating the cloning strategies used to obtain the pAA54b2-3, pAA55b1-1, pYC7A6 and pYC8A6.

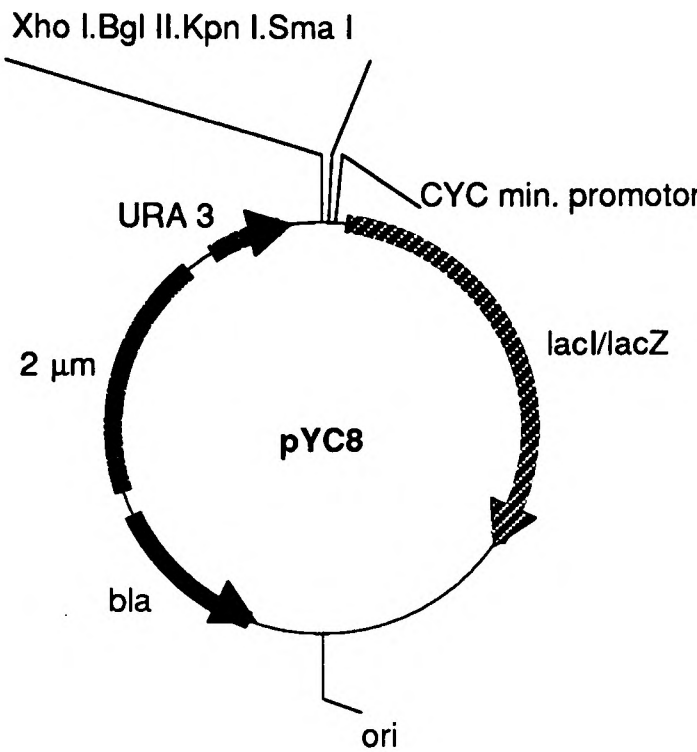


Figure 7.7: Map of the reporter plasmid pYC8

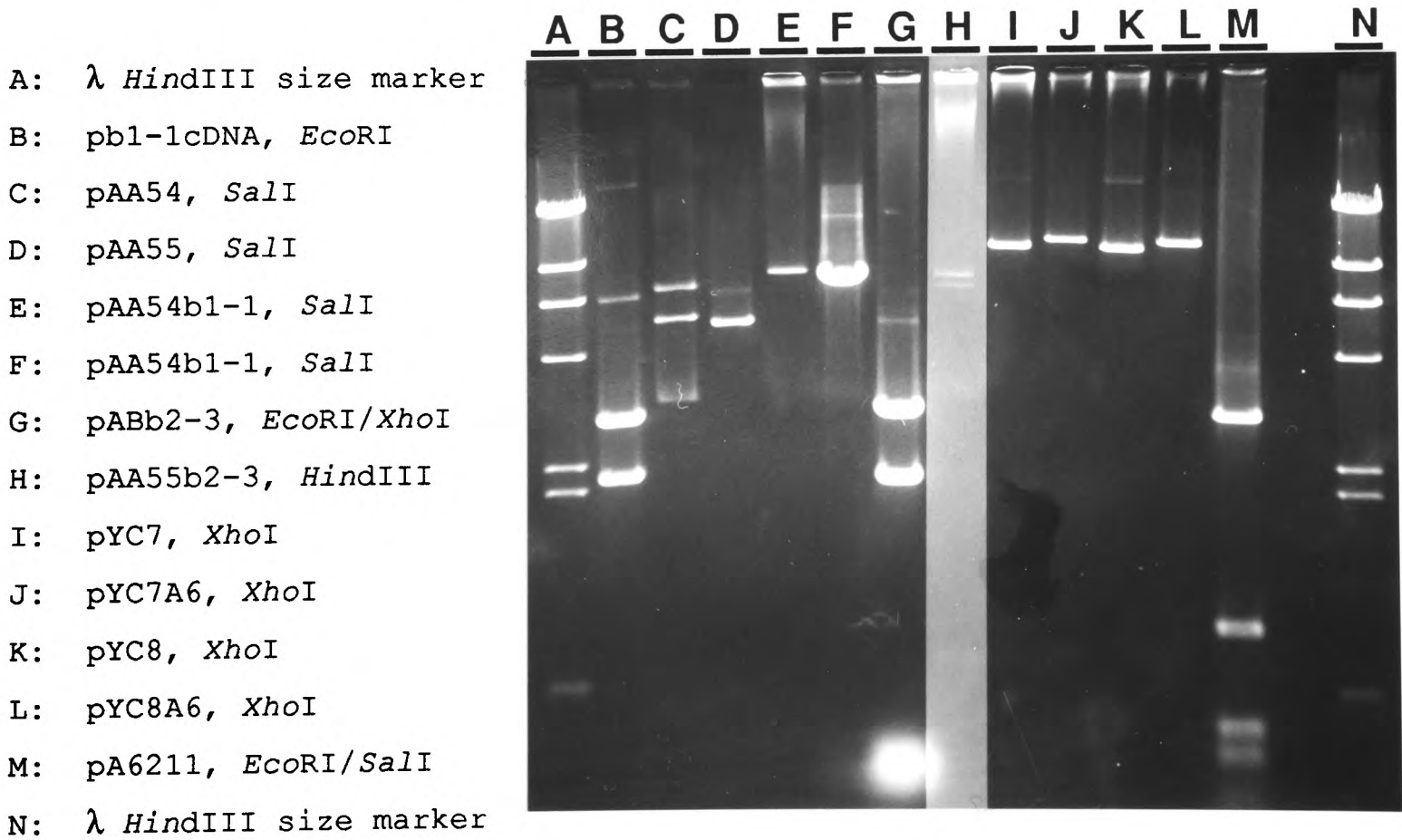


Figure 7.8: Ethidium stained agarose gel showing restriction digests illustrating the cloning strategies for the preparation of pAA54b2-3, pAA55b1-1, pYC7A6 and pYC8A6.

7.2 Transformation of *S. cerevisiae* with the *C. cinereus* mating type genes

Initially as a control, pAA54, pAA55, pYC7 and pYC8 were transformed into *S. cerevisiae* strain YPH500 to show that these were inactive in promoting transcription of the reporter gene. In addition, pAA54/pAA55/pYC7 and pAA54/pAA55/pYC8 were cotransformed into YPH500. The number of transformants obtained in each experiment averaged three for triple and 50 for single transformations. Transformants were replica plated onto galactose medium containing X-Gal and incubated at 30°C for 5-7 days. As expected, none of these control experiments yielded any blue colonies. The experiments were carried out in parallel with transactivation studies by Dr. Banham, showing that failure of colour development was not due to problems with the growth media. The strain YPH500 and plasmids pAA54/55 and pYC7/8 were a gift from A. Andrianopoulos (University of Georgia, Athens, GA) and had been successfully used to demonstrate operator binding by regulatory proteins of *Aspergillus nidulans* (Chang and Timberlake 1992).

pAA54b2-3, pAA55b1-1, pYC7A6 and pYC8A6 were transformed into YPH500 individually and together by cotransformation to give pAA55b2-3/pAA54b1-1/pYC7A6 and pAA55b2-3/pAA54b1-1/pYC8A6. Transformants were replica plated onto galactose medium containing X-Gal and incubated for 5-7 days at 30°C. No blue colonies were obtained in any of these transformations.

There are several possible explanations for this negative result. Assuming that the mating type proteins regulate their own operators in *C. cinereus*, the experiment presented here may not have worked because of potential instability of the *C. cinereus* proteins in yeast. Since the b1-1 protein had been shown earlier (Banham 1993) to function as a strong transcriptional

activator in *S. cerevisiae*, it can be assumed that it should be fairly stable. However this might not be the case for the b2-3 protein. Another possibility could be that the *C. cinereus* proteins are not folded properly after expression in *S. cerevisiae* and cannot activate transcription from their own promoters. Another protein may be required in *C. cinereus* to allow proper binding of any one of the A proteins to an operator sequence in order to activate the transcription machinery. It is also possible that the nuclear localisation signals of the *C. cinereus* proteins are not properly recognised by *S. cerevisiae*, thereby preventing entry into the nucleus. Finally, a post-translational modification necessary for function may not occur in the heterologous yeast system.

More recent results obtained by (Pardo 1993) suggest a much simpler explanation for these negative results in the yeast experiments. The mutant strain NA2 (see beginning of chapter 6) has no active A mating type genes, hence designated the A null mutant. It shows typical monokaryotic morphology. The A6 mutant gene mentioned in the introduction is an in frame fusion of the 5' half of the *a2-1* gene to the 3' half of the *d1-1* gene. The resulting fusion protein contains the HD2 homeodomain of the *a2-1* protein and the potential transactivation domains of the *d1-1* protein, covalently fused together. The presence of this gene results in self-compatibility, that is a constitutive expression of the A regulated pathway of sexual development. When this gene was introduced into the NA2 null mutant by DNA mediated transformation, the developmental switch from monokaryon to clamp cell formation took place.

This result contradicts the initial model (see introduction to this chapter), which postulated that mating type genes control their own rate of transcription. This seems now unlikely, as the mutant gene obviously was

transcribed after DNA mediated transformation, as it exerted its effect in switching the phenotype from monokaryon to clamp cell formation.

It seems therefore, that the experiments described in this chapter are in line with the recent results from transformation studies of the NA2 null mutant using the A6 mutant gene, indicating that the *C. cinereus* mating type genes do not activate their own transcription. It must be said however, that the experiments presented in this chapter should not be over interpreted until expression of the mating type proteins of *C. cinereus* in *S. cerevisiae* has been proved by immunological techniques. Also, the A6 mutant gene may have quite different properties compared with a b1-1/b2-3 heterodimer. It resembles the oncogenes present in viruses of higher eukaryotes in that it is constitutively active. The covalent fusion of the two proteins may result in a folding pattern that renders the mutant protein independent of a third auxiliary factor, that is required for activity of the b1-1/b2-3 heterodimer. A similar situation is found in the truncation of the epidermal growth factor receptor in the v-*erb-B* protein (from avian erythroblastosis virus) which as a result is active permanently (Stoscheck and King 1986).

Experiments in the laboratory suggest that indeed a very special conformation may be required for activity of the A6 mutant fusion protein, because attempts to mimic it by constructing fusion genes *in vitro* have so far not yielded any genes with mutant activity (= self compatibility) (Asante-Owusu 1993). Whether the special conformation of the A6 mutant protein allows it to activate transcription in the experimental set-up described in this chapter could not be analysed, because no cDNA clone for this gene, necessary for the construction of an expression construct, is available. It can therefore not be excluded that the A6 mutant protein can activate its own rate of transcription independently of a third factor, which is

necessary for activity of the b1-1/b2-3 heterodimer and was lacking in the experiments described above.

Chapter 8

Discussion

This study has focused on two genes that comprise part of the *A* mating type factor of the basidiomycete *Coprinus cinereus*. The two genes *b1* and *b2* have multiple allelic forms and encode proteins with distinctly different homeodomain motifs. Different alleles of *b1* and *b2* are sufficient for a compatible *A* mating as shown by the analysis of the A42 and A6 mating type factors. The work described in this study was aimed at trying to understand how this compatible reaction between the *b* genes of A42 and A6 is recognised.

8.1 Gene-pair organisation

To initiate the study of interactions of the *b* proteins from the *C. cinereus* A42 and A6 factors, the *b1-3* and *b2-3* genes from A6 were isolated. DNA sequence analysis of genomic and cDNA clones revealed that the *b* gene-pair of A6 is a divergently transcribed pair of specificity genes (*b1-3* and *b2-3*) analogous to the *b* genes in A42 (*b1-1* and *b2-1*). Sequence comparisons with the *b* gene-pair from A42 demonstrated that the intergenic sequence between *b1* and *b2* is the most dissimilar sequence between the two allelic gene-pairs (35% identity in the intergenic region compared to 55% in the regions coding for the *b* genes). The low degree of DNA sequence similarity in the intergenic region of the *b* genes of A42 and A6 will probably act to prevent recombination which would bring together compatible HD1 and HD2 genes in a single, then self fertile monokaryon.

8.2 Conserved features of b1 and b2 proteins

The b1-3 and b2-3 proteins contain the same two classes of homeodomain motifs (HD1 and HD2) that have been found in the A42 b proteins. The homeodomain is the most conserved region between allelic proteins. This finding supports the idea that this region serves as a DNA binding domain of the A proteins. It is assumed that the A proteins regulate sexual development by acting as transcription factors, i.e. bind to DNA sequences upstream of target genes. The DNA target sites upstream of the regulated genes have to be recognised by proteins of all specificity genes, as these genes are functionally redundant. One would therefore assume that the DNA binding function should be mediated by a very conserved domain in the A proteins since it has been shown that the ability of proteins to bind to identical DNA target sites depends on a conserved DNA binding domain (Pabo and Sauer 1992).

This observation is especially interesting in terms of the functional significance of the HD1 homeodomain found in b1-1 and b1-3. The HD1 domain is an atypical homeodomain as it has an amino acid replacement at one of the four most conserved positions in typical homeodomain sequences. An asparagine has been replaced by an aspartic acid residue which with its carboxyl side chain group cannot mediate the DNA binding function that has been attributed to the amido group of the asparagine in the consensus sequence (Kissinger, Liu et al. 1990) (Wolberger, Vershon et al. 1991). However, predictions of the secondary structure of the HD1 domain (Chapter 4) show that it may have a folding pattern very similar to classical homeodomains. This finding together with the high degree of conservation of the HD1 domain between b1-1 and b1-3 suggests that it is reasonable to assume that the HD1 domain is the conserved DNA binding domain of the HD1 transcription factors.

It is of particular interest to understand how the proteins of these genes can perform absolutely conserved functions (bind to the same DNA target sites, interact with the same general transcription factors, etc.) despite the obvious degree of primary structure divergence. If one assumes that the DNA binding function of both HD1 and HD2 proteins is mediated by the very conserved homeodomain sequences then this conserved function is readily explained. Some other conserved regions of the proteins have been implicated in conserved functions, nuclear localisation signals or PEST sequences (see chapter 4). The C-terminal region of the *C. cinereus* HD1 proteins is conserved (b1-3/b1-1: 73% identity, 83% similarity; b1-3/d1-1: 57% identity, 71% similarity) and shows a high degree of similarity to the C-terminus of the *S. commune* HD1 proteins A α Z3 and Z4 (b1-3/Z3: 31% identity, 42% similarity). Secondary structure predictions for this region suggest an α helical structure. Figure 8.1 shows an alignment of this region of the *C. cinereus* and *S. commune* HD1 proteins. Considering the high level of sequence conservation in this region, it is somewhat surprising, that it can be removed from all HD1 proteins without loss of activity in transformation assays (see chapter 4 and (Tymon, Kües et al. 1992)(Specht, Stankis et al. 1992)). It is possible that this domain of the proteins has an essential function in the monokaryon, but is not required for activity in the clamp cell pathway.

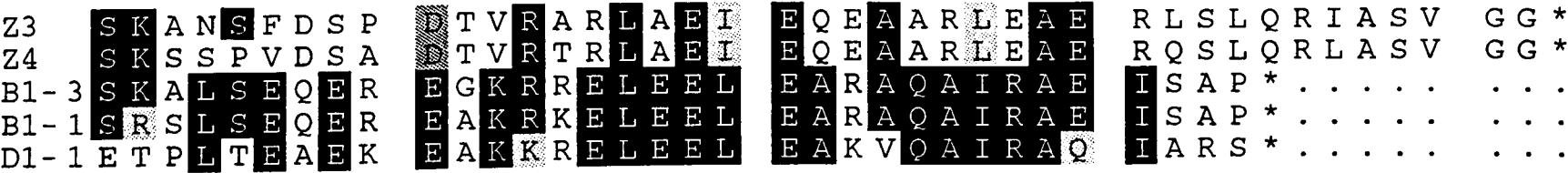


Figure 8.1: Alignment of the conserved C-terminus of *S. commune* and *C. cinereus* HD1 proteins

The proposed transactivation domains in the C-terminal half of the b1-1 protein (Tymon, Kües et al. 1992) are well conserved between b1-1 and b1-3, especially with respect to the abundance of serine, threonine and proline residues implicated in transactivation potential (Mermoud, O'Neill et al. 1989) (Struhl, Struhl et al. 1989) (Ingraham, Flynn et al. 1990) (Tanaka and Herr 1990) (Rambaldi, Nagy Kovacs et al. 1994):

- a) b1-1 327-383: 29% S+T, 19% P
 b1-3 323-376: 28% S+T, 15% P
- b) b1-1 435-496: 16% S+T, 23% P
 b1-3 427-487: 15% S+T, 22% P
- c) b1-1 499-592: 30% S+T, 14% P
 b1-3 490-590: 21% S+T, 12% P

The high degree of conservation between b1-1 and b1-3 in these regions, especially with respect to the abundance of these particular amino acids may be an indication that these regions of the b1 proteins act as transactivation domains in order to stimulate transcription of target genes.

It is more difficult to understand how the most dissimilar regions of both HD1 and HD2 proteins can carry out some conserved functions. It has been shown that the N-termini determine gene-pair specificity (Kües 1993). This means that the quite dissimilar N-terminal regions of b2-1 and b2-3 must possess some sufficiently conserved features that identify them as HD2 proteins of the *b* gene-pair and prevent any positive interaction with HD1 proteins other than b1s. The same must be true for b1-1 and b1-3 in order to prevent them from positively interacting with a2, c2 or d2 proteins. Sequence comparisons of allelic proteins and proteins from different gene-pairs show that the degree of sequence dissimilarity in the N-terminal region up to the homeodomain is considerably higher between proteins from different gene-pairs (b2-3/b2-1: 38% identity/53% similarity; b2-3/a2-1: 19% identity/44% similarity. b1-3/b1-1: 41.3% identity/61.2% similarity; b1-

3/d1-1: 27.4% identity/50.4% similarity). The higher degree of similarity between allelic proteins can therefore be assumed to be a reflection of conserved features responsible for gene-pair specificity.

8.3 The N-terminus of the b proteins determines allele specificity

Experiments with chimeric *b* genes (chapter 4) demonstrated that the N-terminus of both b1 and b2 proteins is responsible for allele specificity. It is thus the least conserved region of allelic proteins that confers specificity. It has been found that allele specificity for the bE and bW proteins from the *b* mating type locus of the hemibasidiomycete *Ustilago maydis* also resides in the N-terminal portion of both HD1 and HD2 proteins (Gillissen, Bergemann et al. 1992). The degree of sequence homologies between the *U. maydis* *b* genes and the *C. cinereus* *A* mating type genes suggests a common evolutionary origin (Kües and Casselton 1992). The four gene-pairs of the *C. cinereus* *A* factor are likely to have evolved by gene-pair duplication and diversification from a simple one gene-pair system. With a common evolutionary origin of the *U. maydis* and *C. cinereus* proteins, it is probable that the way allele specificity is determined may have been at least partly conserved. In this context it is interesting to notice that although the N-terminal sequences of *U. maydis* and *C. cinereus* HD1 and HD2 proteins are not well conserved in primary sequence, the overall spacing appears to be very similar (bE1: 114 amino acids N-terminal of the HD1 domain, b1-3: 116 amino acids N-terminal of the HD1 domain; bW1: 136 amino acids N-terminal of the HD2 domain, b2-3: 157 amino acids N-terminal of the HD2 domain). Also, secondary structure predictions suggest an overall helical structure for this region in HD1 and HD2 proteins from both organisms.

The construction of chimeric alleles of the *bE1* and *bE2* genes has identified the region between amino acids 39 and 87 as a specificity domain (Yee and

Kronstad 1993). Forced recombination in this region leads to chimeric genes with novel specificities. This region and the corresponding region in the *C. cinereus* proteins is predicted to be helical. Helical regions (especially amphipathic helices) have been widely implicated in protein dimerisation (O'Shea, Rutkowski et al. 1989) (O'Shea, Safford III et al. 1989) (O'Shea, Klemm et al. 1991) (Voronova and Baltimore 1990) (Baxevanis and Vinson 1993), often through the formation of coiled coil structures (Crick 1953). It is therefore attractive to think that this region of the HD1 proteins provides the interface for allele recognition, and that the interaction of compatible HD1 and HD2 proteins consists of an HD1/HD2 dimerisation mediated by this region. Two models have been proposed suggesting a mechanism in order to achieve specificity. In the first model (Gillissen, Bergemann et al. 1992), the inactive HD1/HD2 pair dimerises in the unmated cell in such a way, that the DNA binding domains are hidden. Because of the structural differences between allelic forms of the proteins, heterodimerisation then creates a protein in which the homeodomains are available to bind DNA. The second model (Kües and Casselton 1992) proposes that the two different types of homeodomain proteins (HD1 and HD2) can transiently bind adjacent target sites on DNA, but only a compatible HD1/HD2 pair can dimerise in order to stabilise DNA binding and regulate the transcription of target genes. The first model thus suggests that incompatible HD1/HD2 proteins specifically dimerise leading to inactive complexes, whereas the second model suggests that only compatible HD1/HD2 combinations have the ability to dimerise.

8.4 The N-terminus of b1 and b2 proteins contains a dimerisation domain

The unicellular ascomycete *S. cerevisiae* provides a model system for how homeodomain proteins regulate mating compatibility and sexual development. *S. cerevisiae* has two mating types, **a** and α . α -cells produce

the MAT α_1 and MAT α_2 protein, a-cells MATa₁ and MATa₂ (Astell, Ahlstrom-Jonasson et al. 1981)(Nasmyth and Shore 1987). The MAT α_2 and MATa₁ proteins contain homeodomain motifs that have been proposed to be analogous to the HD1 and HD2 type homeodomains found in the basidiomycete mating type proteins (Kües and Casselton 1992). In α haploid cells, α_2 homodimers in conjunction with the general transcription factor MCM1 repress the transcription of a-specific genes (Smith and Johnson 1992) and after mating, the a₁ and α_2 proteins are brought together and repress haploid specific genes as a heterodimer thereby initiating sexual development (Dranginis 1990). It has been shown that a₁ and α_2 binding upstream of haploid specific genes occurs with high cooperativity. The heterodimer shows a much higher sequence specificity for DNA binding and both a₁ and α_2 can be truncated to 85 amino acids (containing the 60 amino acids homeodomain) without losing the ability to cooperatively bind their DNA target site as a heterodimer (Goutte and Johnson 1993). It was concluded that dimerisation through the homeodomain or immediately adjacent regions results in a fixed configuration of the two homeodomains that allows highly specific recognition of operator sequences upstream of haploid specific genes (Goutte and Johnson 1993). A molecular analysis of the α_2 homodimer has shown that homodimerisation is mediated by different N-terminal domains of α_2 and that the correct spacing of the homeodomains is assured through the interaction with MCM1 (Smith and Johnson 1992).

Genetical data from gene replacement experiments with the basidiomycetes *Ustilago maydis* and *C. cinereus* demonstrated that a compatible mating was achieved through an interaction of HD1 and HD2 genes of allelic gene pairs (Gillissen, Bergemann et al. 1992)(Pardo 1993). It was assumed that this HD1/HD2 interaction was analogous to the a₁/ α_2 interaction in *S. cerevisiae*, i.e. formation of heterodimers of homeodomain

transcription factors promotes sexual development (Kües and Casselton 1992). However, in *U. maydis* and in *C. cinereus*, HD1 and HD2 proteins are already present in the haploid cell before mating and a high degree of molecular recognition is required to distinguish compatible from incompatible HD1/HD2 interactions. The current study demonstrates allele specific protein-protein interactions of basidiomycete mating type proteins for the first time (chapter 6). The region of the HD1 proteins responsible was shown to reside in the region N-terminal to the homeodomain, i.e. the first 116 amino acids. The same region of the HD1 proteins is implicated in allele specificity as derived from experiments with chimeric genes (chapter 4) and therefore the observed interaction is likely to be part of the allele recognition process, especially as it was shown that the HD1/HD2 protein interactions were allele specific (chapter 6). b1-1 interacted with b2-3 but failed to interact with b2-1.

Studies with chimeric *bW* genes of *U. maydis* have not been carried out, and it is only known that allele specificity resides in the region N-terminal to the HD2 homeodomain (the first 136 amino acids) (Gillissen, Bergemann et al. 1992). The experiments carried out with chimeric genes of the *C. cinereus* HD2 proteins (Chapter 4) demonstrate that the corresponding region of the *C. cinereus* HD2 proteins is responsible for allele specificity (the first 157 amino acids). The GST association experiments (chapter 6) narrowed down the region essential for the proposed dimerisation of HD1 and HD2 proteins. If dimerisation is the key event in allele recognition then this region should be involved in conferring allele specificity. The experiments demonstrated that the region from amino acids 24 to 115 of the b2-3 protein is sufficient for the interaction with b1-1, and that this specific protein-protein interaction is lost after a further truncation of b2-3 leaving just the first 73 amino acids. This result suggests that the region between

amino acids 73 and 115 of the b2-3 protein contains at least part of the domain, that is responsible for the interaction with the b1-1 protein.

No mating type specific function analogous to the role of the *S. cerevisiae* α_2 homodimer would be expected in basidiomycetes since all haploid cells contain functionally equivalent homeodomain proteins. Homodimers of these proteins should only act as transcription factors if their activity was not correlated to mating. However, inactivation of the A mating type factor in *C. cinereus* leads to a viable phenotype displaying typical monokaryon morphology (Pardo 1993) suggesting that if there is a function for homodimers, it is not essential. The GST association experiments (chapter 6) failed to detect protein-protein interactions between *in vitro* translated b1-1 protein and a truncated b1-1 fusion protein, suggesting that if homodimers analogous to MAT α_2 are formed, the domains involved may be in the C-terminal half of b1-1 which was missing in the fusion protein or that a third protein is required for stable dimer formation.

8.5 Analysis of the N-terminal domains of b1 and b2 proteins

The GST association experiments (chapter 6) demonstrated that the region from amino acids 24 to 115 in b2-3 is sufficient for the interaction with the b1-1 protein. An analysis of the hydrophobic moment (figure 8.3) reveals the possibility of an amphipathic helix from amino acids 57 to 71 in b2-3 and 52 to 66 in b2-1. The hydrophobic moment is a value for the formation of amphipathic structures in protein sequences (Eisenberg, Weiss et al. 1984) (Eisenberg, Schwarz et al. 1984). If it is calculated for an angle of 100° for each peptide bond, it represents a measure for the formation of amphipathic helices. An amphipathic helix at this position could serve as a dimerisation interface for the interaction with b1 proteins in a similar

fashion to the dimerisation of bHLH and bHLHZip transcription factors (Sun and Baltimore 1991) (O'Shea, Rutkowski et al. 1992).

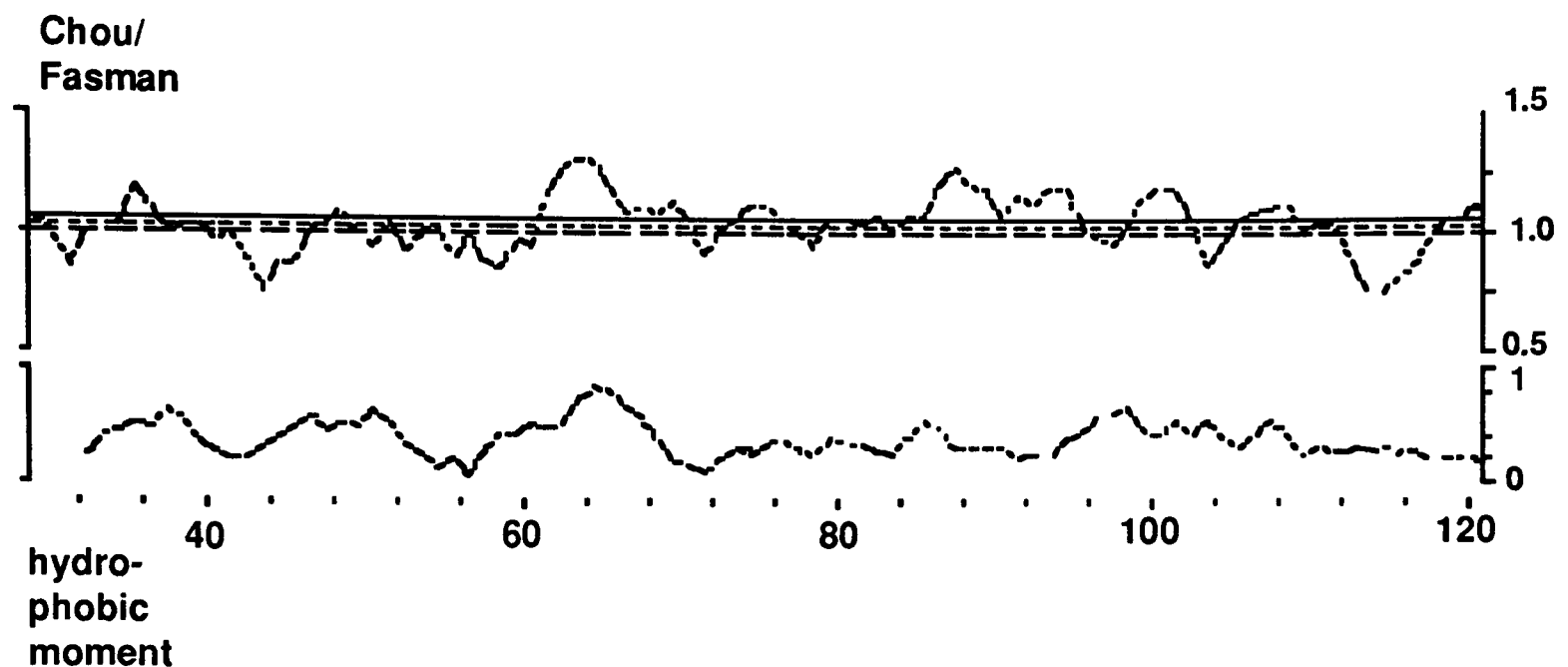


Figure 8.2: Chou/Fasman prediction of probability of α helix formation and hydrophobic moment (calculated for an angle of 100° per peptide bond = α helical structure) of the N-terminal region of the b2-3 protein (amino acids 30 to 120)

Both b2-3 and b2-1 have hydrophobic residues at the (a) and (d) positions of possible heptad repeats suggesting that this region may function in forming a coiled coil structure. Figure 8.3 shows a helical wheel representation of these predicted helices in b2-3 and b2-1 illustrating how they might be involved in the formation of homodimers and heterodimers of HD2 proteins. It is interesting to note that the hydrophobic core of possible dimeric associations is formed by conserved staggered leucine and phenylalanine residues which would perhaps allow the two phenylalanine benzene rings of opposing helices to be stacked on top of each other. Thus their hydrophobic side chains would be buried in the hydrophobic core allowing attracting interactions between the two benzene rings similar to the ones found in the graphite crystal or the stacking forces that are implicated in the stability of the DNA double helix (Saenger 1984). The hydrophilic faces of these proposed helices in b2-3 and b2-1 are not well

conserved (only one out of ten amino acids identical) and since they fall into the region implicated in allele specific dimerisation of HD1 and HD2 proteins these differences may reflect features conferring allele specific interactions. The formation of HD2 homodimers was not analysed using the GST-association assay in this study. It is possible though that *in vivo* HD2 homodimers serve as a means of sequestering reactive amphipathic helices until compatible HD1 proteins are present in order to form HD1/HD2 heterodimers.

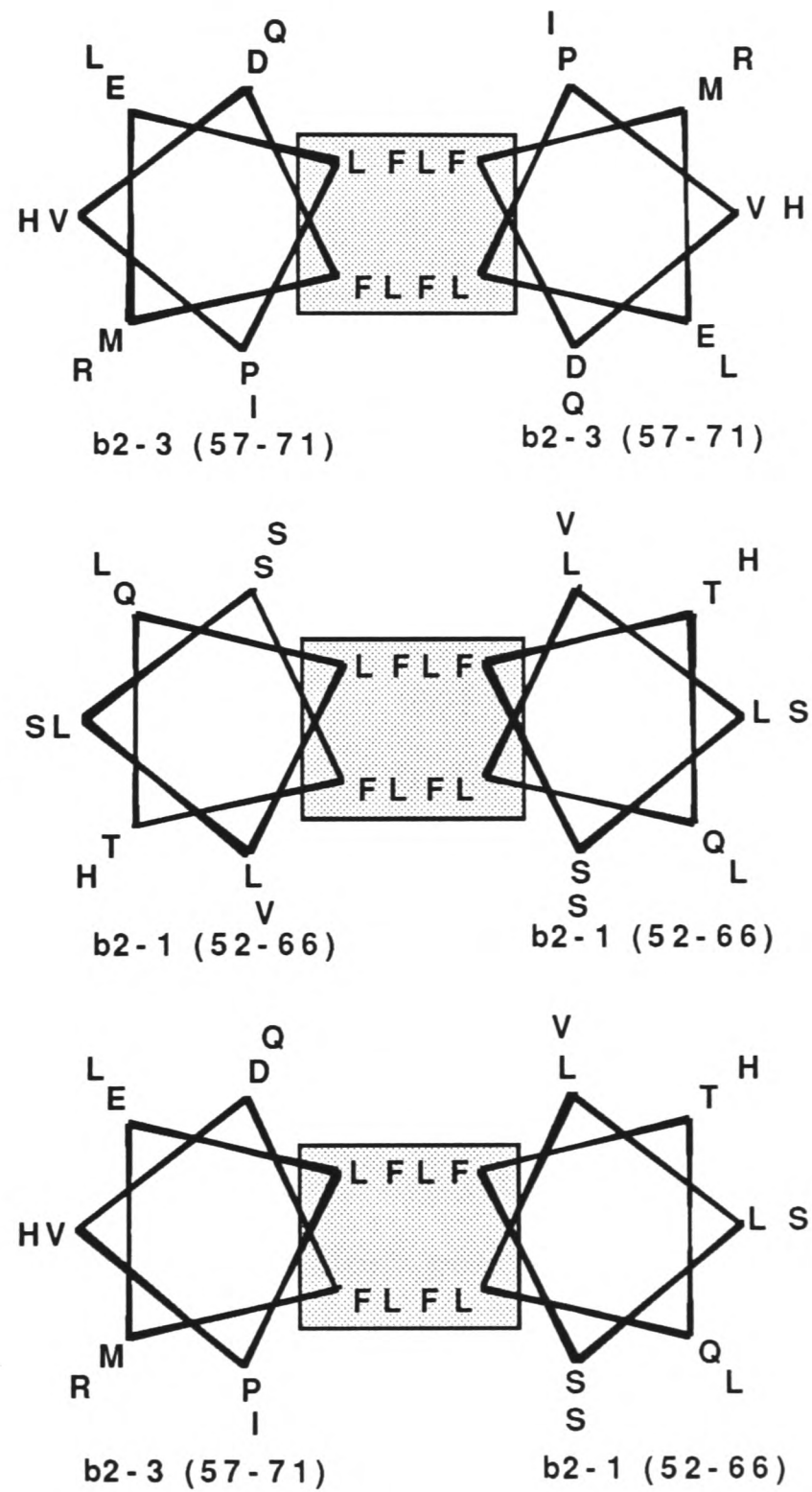


Figure 8.3: Helical wheel presentation of proposed amphipathic helices in b2-3 and b2-1 showing possible b2-3 and b2-1 homodimers and b2-3/b2-1 heterodimers formed in this region (the hydrophobic core of possible helix dimers is highlighted by a shaded box)

The presence of a proline residue in the helix predicted here for b2-3 may not necessarily prevent the formation of an α helix in this region as recent analysis of protein structures has shown that proline residues can occur in α helices. Also, dimerisation may help to stabilise or extend the α helical structure which may not be taken up by the monomer as found for some other transcription factors (Goutte and Johnson 1993) (Ferre-D'Amare, Prendergast et al. 1993).

The work described in this thesis demonstrates that the N-terminal 116 amino acids of the b1-1 protein are sufficient for an allele specific interaction with b2-3. The construction of chimeric alleles of the *U. maydis* *bE1* and *bE2* genes has identified the region between amino acids 39 and 87 as a specificity domain (Yee and Kronstad 1993). Assuming that specificity is achieved by dimerisation, the dimerisation domain of HD1 proteins should be part of this specificity domain. An alignment of the N-terminal sequences of HD1 proteins from *U. maydis* and *C. cinereus* shows that an amphipathic helix may be present in this region. The building block of amphipathic helices is a heptad repeat with hydrophobic residues at the (a) and (d) positions forming the hydrophobic face of the helix (Oas and Endow 1994). As shown in an alignment of parts of the N-termini of *C. cinereus* and *U. maydis* HD1 proteins (figure 8.4), hydrophobic residues are found at positions (a) and (d) of a putative heptad repeat in all *U. maydis* and *C. cinereus* HD1 proteins at equivalent positions, even though there is little sequence conservation otherwise.

b1-1	72	SES	L L A L S E N	M A L L E T K	88
b1-3	72	VAP	L S T L S S N	M V E L D A K	88
d1-1	69	ASL	I A F L P S G	I H Q A E A E	85
bE1	66	THR	I R A V A Q A	F I R F D Q K	82
bE4	66	THR	I N V A V K A	F I C I D Q T	82
bE6	66	AHR	L E V A V K V	F L H I D R K	82
			a d a d		

Figure 8.4: Putative heptad repeat with hydrophobic residues at positions (a) and (d) in the N-terminus of *C. cinereus* and *U. maydis* HD1 proteins

A helical wheel representation of the two predicted amphipathic helices in b1-1 and b1-3 is shown in figure 8.5. The hydrophobic face of the helix and neighbouring residues are conserved, whereas the centre of the hydrophilic face is dissimilar between the two helices (ALET LA in b1-1 and TESASV in b1-3) as for the predicted amphipathic helix in the b2 proteins. Again, the observation that dissimilarities between the two allelic proteins are confined to one face of a putative amphipathic helix suggests that these variable regions may be involved in conferring allele specificity. The conservation of the residues on the hydrophobic face of these amphipathic helices is necessary in order to maintain the proposed mechanism of dimerisation, i.e. association of two amphipathic helices. Figure 8.5 also shows how b1 proteins could possibly form dimers mediated by this region. The GST-association experiments (chapter 6) did not show any evidence for b1-1 homodimer formation. The length of the helices shown in figure 8.5 (14 amino acids long) suggests that this helix alone would not be able to mediate the formation of stable leucine zipper like dimers. A typical leucine zipper (28 amino acids long (Landschultz, Johnson et al. 1988) (O'Shea, Rutkowski et al. 1989)) is twice as long as the amphipathic helices predicted to be present in the N-terminus of b1 and b2 proteins. The possible formation of b1 homodimers could be achieved through the interaction with a third protein. It has been found for the similar HNF1 protein which also contains an atypical homeodomain that homodimer formation is assisted

by a small protein called "dimerisation cofactor of HNF1 = DCoH) (Mendel, Khavari et al. 1991). None of the dimerisation assays used in this study allowed for the possible requirement of a third protein in order to detect b1 homodimers.

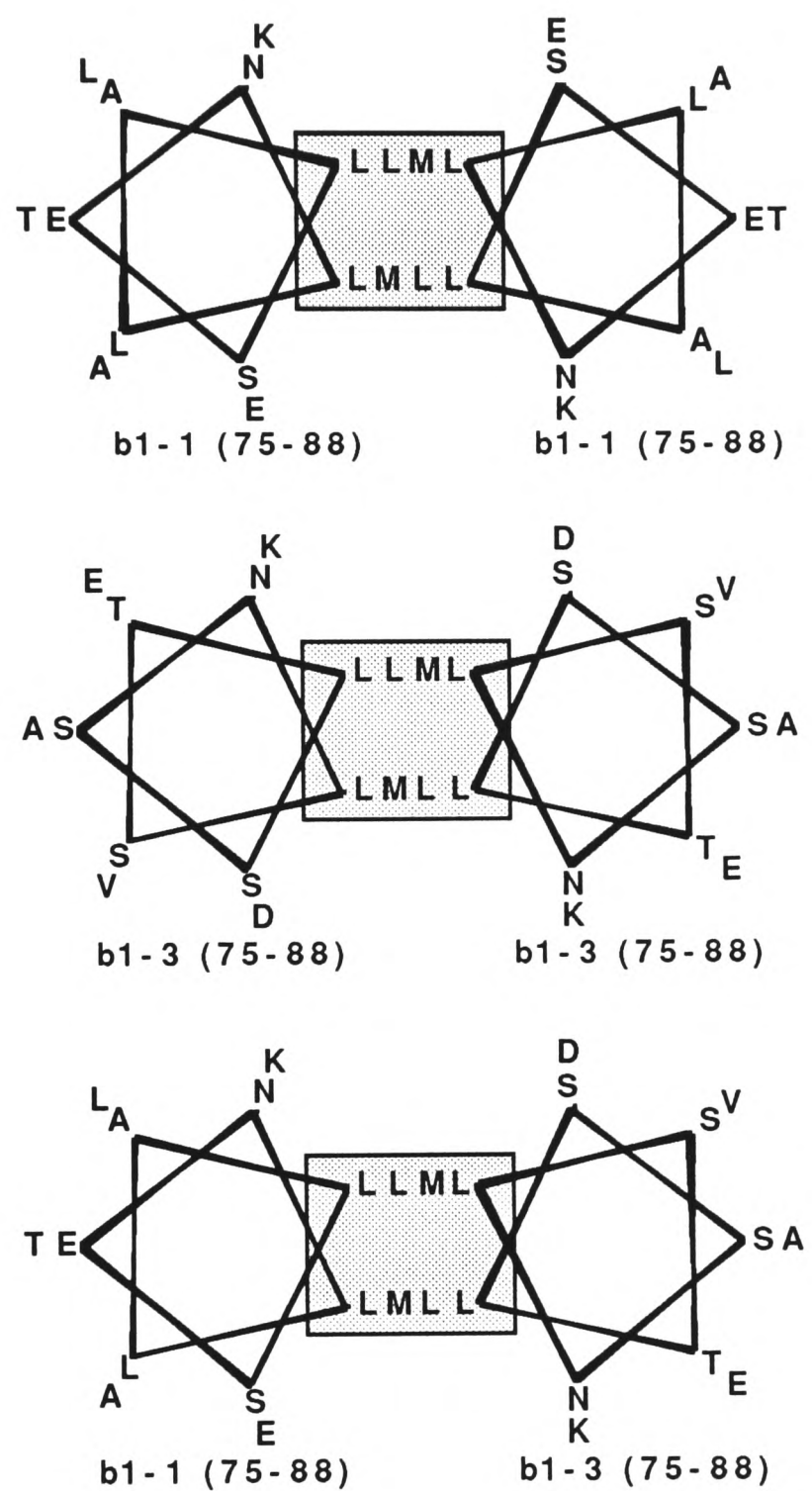


Figure 8.5: Predicted amphipathic helices in the N-termini of b1-1 with possible formation of HD1 protein dimers mediated by this region (the hydrophobic core of possible protein-protein associations is highlighted by a shaded box)

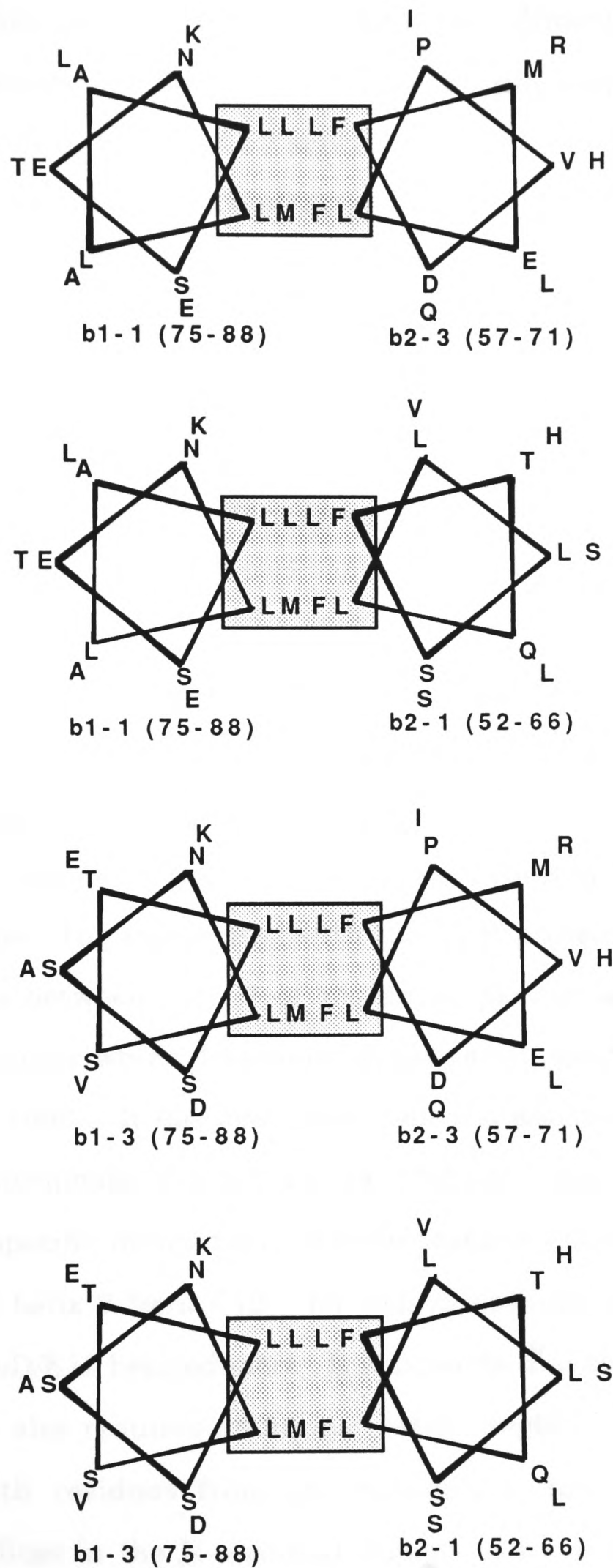


Figure 8.6: Formation of HD1/HD2 heterodimers mediated by the amphipathic helices in the N-terminal regions of b1 and b2 proteins.

The amphipathic helices that mediate the dimerisation of HLH transcription factors are similar in length to the amphipathic helices predicted for the N-terminus of b1 and b2 proteins (14 amino acids for b1 and b2 proteins (see above), 11 and 15 amino acids for the helices 1 and 2 respectively of the HLH protein MyoD (Shirakata, Friedman et al. 1993)). Figure 8.6 shows how the predicted amphipathic helices of b1 and b2 proteins may be involved in the formation of b1/b2 heterodimers. This analysis does not give any clues, however, as to why compatible b1/b2 dimers should be more stable than incompatible ones, as predicted from the results of the GST-association experiments (chapter 6).

Keeping in mind though that the interaction of the HLH proteins requires two helices of similar length to the amphipathic helices predicted for the N-terminus of b1 and b2 proteins from each interaction partner, it is probable that additional regions in the N-terminus are required for allele specific b1/b2 interactions. The comparison with the HLH proteins also shows that it is interactions between helix 1 of one dimer partner and helix 2 of the second dimer partner which determine dimerisation specificity (Shirakata, Friedman et al. 1993). It has been shown for the association of MyoD and E12 that the hydrophobic faces of the two helices 1 and the two helices 2 interact, while specific interactions of hydrophilic residues between helix 1 from MyoD and helix 2 from E12 (and vice versa) lead to the preferential formation of MyoD/E12 heterodimers. It is possible that the interaction of b1 and b2 proteins also requires additional helical regions that allow specific interactions with residues from the hydrophilic face of the predicted amphipathic helices in the N-terminal regions of b1 and b2 proteins. An analysis of the respective protein sequences does not reveal any strong evidence for additional amphipathic helices. However, within the region from amino acids 73 to 115 of the b2-3 protein, which was shown to be essential for the interaction with the b1 protein (chapter 6), secondary

structure predictions implicate a helical region from amino acids 86 to 102. A helical structure is also predicted for the equivalent region in b2-1 from amino acids 81 to 97.

A helical wheel presentation of the two predicted helices shows that the position of all hydrophobic residues is conserved between b2-1 and b2-3. The dissimilarities between the two sequences are clustered on the side facing to the left: QKQE for b2-3 compared to TTER for b2-1 (figure 8.7). It is tempting to speculate that this feature may reflect the function of this region in being part of an allele specific interface. The relatively high degree of structural conservation would allow dimerisation to all b1 proteins (besides the b1 protein coming from the same gene-pair as the respective b2 protein). The variable face of this helix together with the highly variable hydrophilic faces of the predicted amphipathic helices may contain the structural features necessary to prevent for instance the b1-1 protein from dimerising with the b2-1 protein. The results presented in this study therefore for the first time allow speculation about the function of specific regions of the *C. cinereus* A mating type proteins in the molecular recognition process that controls mating compatibility.

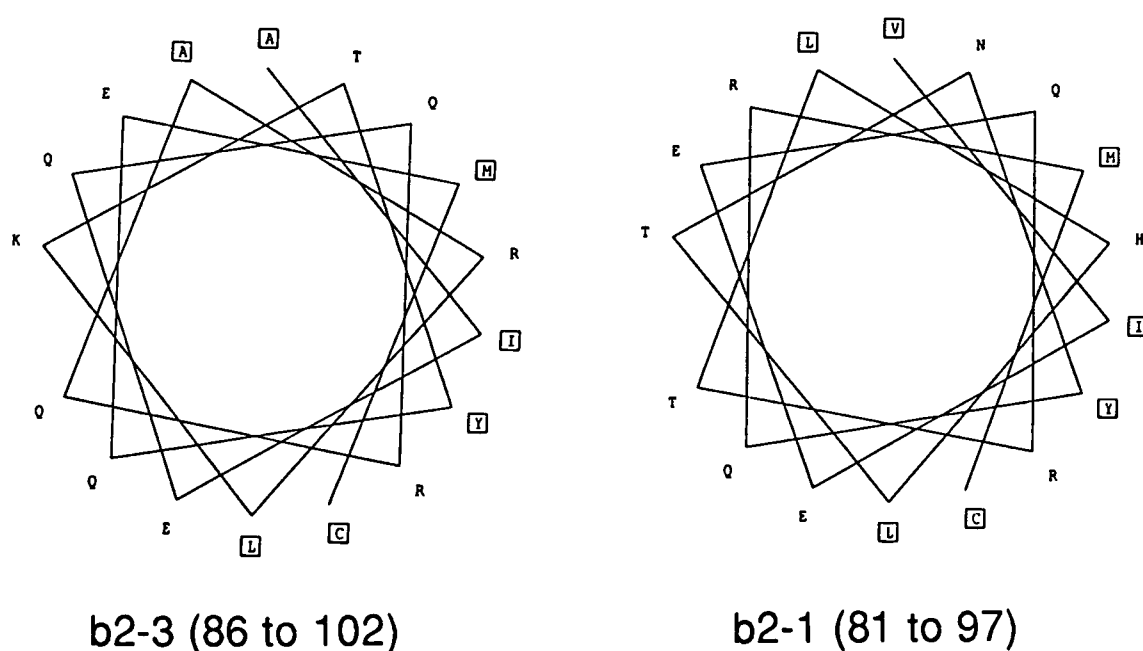


Figure 8.7: Predicted helices in the N-terminal regions of b2-3 and b2-1 (all hydrophobic residues are boxed)

The specific dimerisation of HLH transcription factors is regulated by an equivalent mechanism (Shirakata, Friedman et al. 1993). It has been shown that non-specific dimerisation is mediated by the hydrophobic face of the two amphipathic helices in the helix-loop-helix domain. Myogenic bHLH (basic helix-loop-helix) proteins like MyoD form heterodimers with ubiquitous bHLH proteins, known as E proteins, and activate transcription of muscle-specific genes by binding to the E-box consensus sequence in muscle gene operators (Olson and Klein 1994). In general, all of these HLH proteins can form either homodimers or heterodimers, however MyoD preferably forms a heterodimer with the E12 protein rather than a homodimer (Sun and Baltimore 1991). The implication of these findings is that dimerisation of the bHLH proteins in the cell may be highly competitive and therefore play a major role in gene regulation. Studies of the specific heterodimerisation of MyoD and E12 showed that charged residues from the hydrophilic face of the amphipathic helices provide attractive forces (salt bridges) to allow specific heterodimerisation and repulsive forces to prevent non-specific heterodimerisation (Shirakata, Friedman et al. 1993).

An analogous mechanism has been shown to mediate preferential heterodimerisation of the basic leucine zipper (bZIP) class of DNA binding proteins (O'Shea, Rutkowski et al. 1992) (Vinson, Hai et al. 1993). Again, the hydrophobic face of the leucine zipper (also an amphipathic helix) mediates homo- and heterodimerisation, whereas the charged residues from the hydrophilic faces of for instance the FOS and JUN leucine zippers promote the more specific heterodimerisation, so that it is concluded that the Jun/Fos heterodimer is the predominant form found *in vivo*.

Further evidence for the proposed mechanism of *C. cinereus* mating type protein interactions comes from recently published experimental data (Chi-Yip, Adamson et al. 1994) suggesting that the heterodimerisation of the S.

cerevisiae mating type proteins a_1 and α_2 is not mainly mediated by their C-terminal regions as proposed by (Goutte and Johnson 1993) but also involves the interaction of putative coiled coil forming helices in the N-terminal regions. These new results indicate that the mechanism of homeodomain protein interactions may have been conserved between *C. cinereus* and *S. cerevisiae*. However, the far higher complexity of the *C. cinereus* A factor requires discrimination between compatible and incompatible heterodimer partners which seems to be achieved by a failure of dimerisation between incompatible proteins (see chapter 6).

8.6 Mechanism of a compatible A mating type reaction

The question asked by mycologists for many years is how is self distinguished from non-self in the mating reaction of the basidiomycetes? The results obtained in this study provide a mechanism (protein dimerisation) and suggest, that the molecular recognition event that governs this self/non-self recognition might be more specific in the recognition of self, rather than specifically recognising non-self. This means that, for example, a particular region of the b2 proteins has a conserved helical structure, with variable faces. The conserved features of these helices allow dimerisation of b2 proteins with b1 proteins from different alleles of the same gene-pair. The variable faces contain features that prevent dimerisation of b2 and b1 proteins coming from the same gene pair, i.e. a specific repulsion of incompatible proteins. Two further sets of experiments underline this model:

Firstly, the studies using recombined *bE* genes in *U. maydis* defined a region conferring specificity in the N-terminus of the bE proteins (amino

acids 39 to 87) (Yee and Kronstad 1993). It was found that recombination of two allelic *bE* genes in this region yielded mutant genes with novel *bE* specificities, that could interact with both allelic wild *bW* partner genes (a *bE1/bE2* hybrid could interact with *bW1* and *bW2*). It seems therefore likely that the inability of *bE1* to interact with *bW1* and *bE2* with *bW2* is a specific feature which is lost in the hybrid *bE* genes. These findings underline the necessity that the HD1 and HD2 genes of a gene pair must stay together during evolution, since it has to be postulated that they evolve as a pair and that diversification occurs in parallel, i.e. mutations in the HD1 gene have to be "silenced" by mutations in the HD2 gene from the same gene pair. Otherwise monokaryons with activated A factor function would occur in *C. cinereus*. Such monokaryons might be compromised in nature, however, because they undergo a wasteful process of partial development. Evolution would therefore lead to very specifically matched pairs of non-interacting (incompatible) HD1 and HD2 genes.

Secondly, it has been found that individual *C. cinereus* A mating type genes can promote sexual development after transformation into the heterologous host *Coprinus bilanatus* (Challen, Elliott et al. 1993). This result shows that *C. cinereus* A proteins can interact with the heterologous *C. bilanatus* mating type proteins, again suggesting that the recognition of a compatible protein may not be very specific, but that it is the repulsion of incompatible proteins from the same gene-pair that requires higher specificity.

The level of complexity of the interactions of the A mating type proteins of *C. cinereus* is higher than that of the genes of the *U. maydis* b mating type locus. The *C. cinereus* A factor is composed of functionally redundant pairs of HD1/HD2 specificity genes as opposed to the single HD1/HD2 pair present at the *U. maydis* b mating type locus. Molecular analysis of several A factors showed that only allelic HD1/HD2 combinations can promote

sexual development (Kües and Casselton 1993). It therefore has to be postulated that b2-1 for example only interacts with b1 proteins other than b1-1, and does not positively interact with a1, c1 and d1 proteins. It is interesting to note in this context that although the secondary structure prediction for the N-terminal region of the a2-1 protein shows an overall helical structure, the position of predicted helices is shifted when one aligns a2-1 to b2-1 and b2-3. The second helix implicated in allele specific recognition of b2 proteins is from amino acids 86 to 102. a2-1 has a strongly predicted helix from position 69 to 87 (figure 8.6). It is interesting to speculate that this helix could represent part of an interface responsible for the allele specific recognition of a2 proteins and that the different positions of these helices in a2 and b2 proteins prevent any interaction between proteins from different gene-pairs. This would mean that the mechanism of HD1/HD2 interaction using helices as interfaces is conserved, however the spacing of the amphipathic helix in HD2 proteins determines whether interaction occurs with a1, b1, c1 or d1 proteins. The region N-terminal to the HD2 homeodomain is 20 amino acids longer for the *C. cinereus* HD2 proteins (157 amino acids) than for the *U. maydis* bW proteins (136 amino acids) and it is possible that these additional amino acids allow for the proposed different spacing of amphipathic helices in the *C. cinereus* HD2 proteins from different gene-pairs, a requirement not necessary in the one gene-pair system of *U. maydis*.

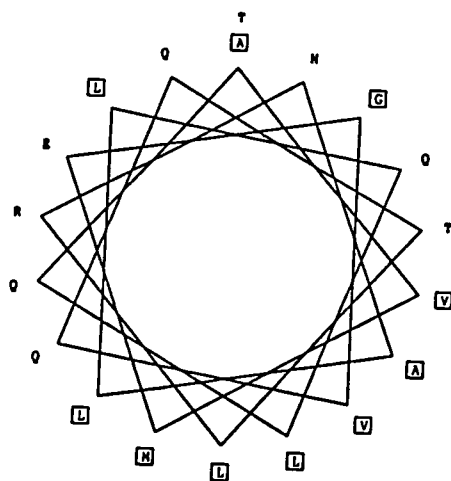


Figure 8.6: Predicted helix in the N terminus of the a2-1 protein (all hydrophobic residues are boxed)

The experiments described in this study have developed a powerful experimental system to study transcription factor interactions in the regulation of the development of a eukaryotic organism. The GST fusion protein *in vitro* association assay allowed the detection of specific protein-protein interactions (chapter 6). These studies identified a short region of the b2 proteins responsible for the specific interaction with b1 proteins, and it is now possible to carry out the reciprocal experiment in order to precisely locate the domain in b1 proteins that mediates the interaction with the b2 proteins. It might then be possible to predict salt bridges that stabilise compatible heterodimers or residues that repel incompatible HD1/HD2 proteins and the significance of specific amino acid residues could be analysed by site directed mutagenesis in analogy to the work carried out with the Jun/Fos and MyoD/E12 heterodimeric transcription factors (O'Shea, Rutkowski et al. 1992) (Shirakata, Friedman et al. 1993). The heterologous expression system for b1-1 and b2-3 in *Saccharomyces cerevisiae* (chapter 7) can be utilised to analyse potential DNA target sites of the *C. cinereus* A mating type proteins, and together with the antibodies raised against the b1-1 protein (chapter 5) and the GST association assay (chapter 6), a detailed molecular study of A mating type protein function becomes possible.

The regulation of development in eukaryotes depends on the specific interactions of transcription factors with operator sequences upstream of target genes. Dimerisation of transcription factors offers enormous scope for regulation (Lamb and McKnight 1991). It has been found that different dimer combinations of bZIP and HLH proteins bind to their target sites with distinctly different affinities, and that the differences in binding reflect differences in dimer stability (Murre, McCaw et al. 1989). It has also been shown that the DNA binding specificity of isolated homeodomain sequences

is low (Desplan, Theis et al. 1988) (Hoey and Levine 1988) (Hayashi and Scott 1990) (Catron, Iler et al. 1993) and that heterodimerisation of the *S. cerevisiae* Mata₁ and Mata₂ proteins allows highly specific binding to target sites that are only bound at very low affinity and specificity by an α_2 homodimer and not at all by α_1 (Goutte and Johnson 1988) (Goutte and Johnson 1993). Some homeodomain proteins seem to achieve higher binding specificity by containing numerous DNA binding motifs. The ATBF1 protein for example has been shown to contain 4 homeodomains and 17 zinc finger motifs (Yasuda, Mizuno et al. 1994). There are however no obvious additional DNA binding motifs in the *C. cinereus* A proteins besides the single HD1 and HD2 homeodomains.

Earlier findings had suggested a potential POU specific domain like region in the N-terminal half of the b1-1 protein (Tymon, Kües et al. 1992). The POU transcription factors belong to a special class of homeodomain proteins and have a bipartite DNA binding domain, composed of a homeodomain and a POU specific domain (Herr, Sturm et al. 1988) (Verrijzer and van der Vliet 1993). The POU specific domain has been shown to confer high specificity in DNA binding (Verrijzer, Alkema et al. 1992). The region in b1-1 found to be analogous to the POU specific domain lies in the N-terminal half of the protein. This is the least conserved region between the allelic proteins b1-1 and b1-3 (N-terminal region: 41% identity, homeodomain: 72% identity, C-terminal region: 67% identity), and it is therefore unlikely that this region should be involved in highly specific DNA binding. It now seems more likely that the specific dimerisation of compatible HD1/HD2 proteins brings together the two classes of homeodomain motifs which can then lead to the stable binding of this heterodimer at operator sites of target genes, maybe with each homeodomain occupying one half site of a bipartite DNA recognition site thereby achieving highly sequence specific binding. The POU specific

domain has been shown to mediate cooperative DNA binding of POU transcription factors leading for example to Pit1/GHF1 dimers bound to dimeric DNA target sequences (Ingraham, Flynn et al. 1990). The sequence similarities between the N-terminus of b1-1 and the POU domain may thus be explained by the fact, that both regions mediate protein-protein interactions and are helical (secondary structure for b1-1 predicted, for POU specific domain see (Dekker, Cox et al. 1993)). The specific protein-protein interactions of the A proteins of *C. cinereus* can now be studied at a molecular level and the analogies to transcription factors of higher eukaryotic organisms underline the importance of a further analysis of A protein function. Specific dimerization of compatible HD1/HD2 proteins can be seen as an activation step bringing together complementary DNA binding and transactivation domains. Similarly, oligomerisation has been identified as the key event in the activation of the human heat shock factor (HSF) and the rat E box binding (REB α and REB β) proteins (de la Brousse and McKnight 1993). The constitutively active A mutant gene (in-frame fusion of an HD1 and an HD2 gene) can be interpreted as an analogy to chromosome translocations leading to transcription factor fusion proteins in higher eukaryotes (Sorensen, Lessnick et al. 1994) (Green 1992) thereby stressing the more general applicability of any molecular mode of action that may be deduced for the A proteins using the experimental system developed in this thesis.

References

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J. Mol. Biol. **215**: 403-410.

Andres, V., M. D. Chiara, et al. (1994). "A new bipartite DNA-binding domain: cooperative interaction between the *cut* repeat and the homeo domain of the *cut* homeo proteins." Genes & Dev. **8**: 245-257.

Appel, B. and S. Sakonju (1993). "Cell-type specific mechanisms of transcriptional repression by the homeotic gene products UBX and ABD-A in *Drosophila* embryos." EMBO J. **12**(3): 1099-1109.

Asante-Owusu, R. (1993). (Personal communication)

Astell, C. R., L. Ahlstrom-Jonasson, et al. (1981). "The sequence of the DNAs coding for the mating-type loci of *Saccharomyces cerevisiae*." Cell **27**: 15-23.

Ballance, D. J. (1986). "Sequences important for gene expression in filamentous fungi." Yeast **2**: 229-236.

Ballance, D. J. (1991). Transformation systems for filamentous fungi and an overview of fungal gene structure. Molecular Industrial Mycology: Systems and Applications for Filamentous Fungi Eds. S. A. Leong and R. M. Berka. New York, Marcel Dekker Inc. 1-29.

Banham, A. (1993). (Personal communication)

Baxevanis, A. D. and C. R. Vinson (1993). "Interactions of coiled coils in transcription factors: where is the specificity?" Curr. Opin. in Genet. & Dev. **3**: 278-285.

Beachy, P. A., J. Varkey, et al. (1993). "Cooperative binding of an *Ultrabithorax* homeodomain protein to nearby and distant DNA sites." Mol. Cell. Biol. **13**: 6941-6956.

Bienz, M. (1994). "Homeotic genes and positional signalling in the *Drosophila* viscera." TIG **10**(1): 22-26.

Binniger, D. M., C. Skrzynia, et al. (1987). "DNA-mediated transformation of the basidiomycete *Coprinus cinereus*." EMBO J. **6**: 835-840.

Blackwell, J. R. and R. Horgan (1991). "A novel strategy for production of a highly expressed recombinant protein in an active form." FEBS Letts. **295**: 10-12.

Budd, P. S. and I. J. Jackson (1991). "Production of mouse Hox-2.1 protein in *Escherichia coli*: characterisation of *in vitro* binding to DNA." Gene **105**: 213-219.

Casselton, L. A. (1978). Dikaryon formation in higher basidiomycetes. The Filamentous Fungi. Eds. J. E. Smith and D. R. Berry. London, Edward Arnold. 275-297.

Casselton, L. A. (1993). (Personal communication)

Casselton, L. A. and A. de la Fuente Herce (1989). "Heterologous gene expression in the basidiomycete fungus *Coprinus cinereus*." Curr. Genet. **16**: 35-40.

Catron, K. M., N. Iler, et al. (1993). "Nucleotides flanking a conserved TAAT core dictate the DNA binding specificity of three murine homeodomain proteins." Mol. & Cell. Biol. **13**: 2354-2365.

Cerenkov, P. (1934). Akad. Nauk SSSR Dokl. **2** . Akad. Nauk SSSR. 451.

Ceska, T. A., M. Lamers, et al. (1993). "The X-ray structure of an atypical homeodomain present in the rat liver transcription factor LFB1/HNF1 and implications for DNA binding." EMBO J. **12**(5): 1805-1810.

Challen, M. P., T. J. Elliott, et al. (1993). "Expression of A mating type genes of *Coprinus cinereus* in a heterologous basidiomycete host." Mol. Gen. Genet. **241**: 474-478.

Chang, Y. C. and W. L. Timberlake (1992). "Identification of *Aspergillus brlA* response elements (BREs) by genetic selection in yeast." Genet. **133**: 29-38.

Chen, P., P. Johnson, et al. (1993). "Multiple ubiquitin-conjugating enzymes participate in the *in vivo* degradation of the yeast MAT α_2 repressor." Cell **74**: 357-364.

Chi-Yip, H., J. G. Adamson, et al. (1994). "Heterodimerization of the yeast MAT α_1 and MAT α_2 proteins is mediated by two leucine zipper-like coiled-coil motifs." EMBO J. **13**(6): 1403-1413.

Chiao, P. J., S. Miyamoto, et al. (1994). "Autoregulation of IKB α activity." Proc. Natl. Acad. Sci. USA **91**: 28-32.

Chien, C.-T., P. L. Bartel, et al. (1991). "The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest." PNAS U.S.A. **88**: 9578-9582.

Chirgwin, J. M., A. E. Przybyla, et al. (1979). "Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease." Biochemistry **18**: 5294-5299.

Crick, F. H. C. (1953). "The packing of α -helices: simple coiled coils." Acta Cryst. **6**: 689-697.

Dagert, M. and S. D. Ehrlich (1979). "Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells." Gene **6**: 23-28.

Day, P. R. (1960). "Mutations affecting the A mating type locus in *Coprinus lagopus*." Genetical Society of Great Britain : 457.

Day, P. R. (1960). "The structure of the A mating type factor in *Coprinus lagopus*." Genetics **45**: 641-650.

Day, P. R. (1963). "The structure of the A mating type factor in *Coprinus lagopus*: wild alleles." Genet. Res. Camb. **4**: 323-325.

de la Brousse, F. C. and S. L. McKnight (1993). "Glimpses of allostery in the control of eukaryotic gene expression." TIG **9**(5): 151-154.

Dekker, N., M. Cox, et al. (1993). "Solution structure of the POU-specific DNA-binding domain of Oct-1." Nature **362**: 852-855.

Desplan, C., J. Theis, et al. (1988). "The sequence specificity of homeodomain-DNA interaction." Cell **54**: 1081-1090.

Dowell, S. J., J. S. H. Tsang, et al. (1992). "The centromere and promoter factor 1 of yeast contains a dimerization domain located carboxy-terminal to the bHLH domain." NAR **20**: 4229-4236.

Dranginis, A. M. (1990). "Binding of yeast $\alpha 1$ and $\alpha 2$ as a heterodimer to the operator DNA of a haploid-specific gene." Nature **347**: 682-685.

Eisenberg, D., E. Schwarz, et al. (1984). "Analysis of membrane and surface proteins with the hydrophobic moment plot." J. Mol. Biol **179**(1): 125-142.

Eisenberg, D., R. M. Weiss, et al. (1984). "The hydrophobic moment detects periodicity in protein hydrophobicity." Proc. Natl Acad. Sci. USA **81**(1): 140-144.

Ferre-D'Amare, A. R., E. B. Prendergast, et al. (1993). "Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain." Nature **363**: 38-45.

Fritz, C. C. and M. R. Green (1992). "Fishing for partners." Curr. Biol. **2**(8): 403-405.

Giasson, L., C. A. Specht, et al. (1989). "Cloning and comparison of the A α mating-type alleles of the basidiomycete *Schizophyllum commune*." Mol. Gen. Genet. **218**: 72-77.

Giesy, R. M. and P. R. Day (1965). "The septal pores of *Coprinus lagopus* (Fr.) *sensu* Buller in relation to nuclear migration." American Journal of Botany **52**: 287-294.

Gillissen, B., J. Bergemann, et al. (1992). "A two-component system for self-non-self recognition in *Ustilago maydis*." Cell **68**: 647-657.

Glisin, V. R., C. Crkvenjakov, et al. (1974). "Ribonucleic acid isolated by caesium chloride centrifugation." Biochemistry **13**: 2633-2637.

Goldfarb, A. N., S. Goneli, et al. (1992). "T-Cell Acute Lymphoblastic Leucemia - the associated gene SCL/tal codes for a 42 kd nuclear phosphoprotein." Blood **80**(11): 2858-2866.

Goutte, C. and A. D. Johnson (1988). "a1 protein alters the DNA binding specificity of the $\alpha 2$ repressor." Cell **52**: 875-882.

Goutte, C. and A. D. Johnson (1993). "Yeast a1 and $\alpha 2$ homeodomain proteins form a DNA-binding activity with properties distinct from those of either protein." J. Mol. Biol. **233**: 359-371.

Green, A. R. (1992). "Transcription factors, translocations and haematological malignancies." Blood Reviews **6**: 118-124.

Group, G. C. (1991). Program Manual for the GCG Package, Version 7, April 1991. Madison, Wisconsin USA 53711, Genetics Computer Group, 575 Science Drive.

Gurr, S. J., S. E. Unkles, et al. (1987). The structure and organization of nuclear genes of filamentous fungi. Gene Structure in Eukaryotic Microbes Ed. J. R. Kinghorn. Oxford, IRL Press Limited. 93-139.

Hanahan, D. (1983). "Studies on transformation of *Escherichia coli* with plasmids." J. Mol. Biol. **166**: 557-58.

Harlow, E. and D. Lane (1988). Antibodies: A Laboratory Manual. Cold Spring Harbor, New York., Cold Spring Harbor Laboratory.

Harrison, S. C. and A. K. Aggarwal (1990). "DNA recognition by proteins with the helix-turn-helix motif." Ann. Rev. Biochem. **59**: 933-969.

Hayashi, S. and M. P. Scott (1990). "What determines the specificity of action of *Drosophila* homeodomain proteins?" Cell **63**: 883-894.

Haylock, R. W., A. Economou, et al. (1980). "Dikaryon formation in *Coprinus cinereus*: Selection and identification of *B* factor mutants." J. Gen. Mic. **121**: 17-26.

Herr, W., R. A. Sturm, et al. (1988). "The POU domain: a large conserved region in the mammalian *pit-1*, *oct-1*, *oct-2* and *Caenorhabditis elegans unc-86* gene products." Genes Dev. **2**: 1513-1516.

Herschbach, B. M. and A. D. Johnson (1993). "The yeast $\alpha 2$ protein can repress transcription by RNA polymerases I and II but not III." Mol. & Cell. Biol. **13**(7): 4029-4038.

Herskowitz, I. (1989). "A regulatory hierarchy for cell specialization in yeast." Nature **342**: 749-757.

Hirai, S. I., H. Kawasaki, et al. (1991). "Degradation of transcription factors c-Jun and c-Fos by calpain." FEBS Lett. **287**: 57-61.

Hoey, T. and M. Levine (1988). "Divergent homeobox proteins recognize similar DNA sequences in *Drosophila*." Nature **332**: 858-861.

Hsu, H. L., J. T. Cheng, et al. (1991). "Enhancer-binding activity of the *tal-1* oncoprotein in association with the E47/E12 helix-loop-helix proteins." Mol. Cell. Biol. **11**: 3037-3042.

Hu, J. C., E. K. O'Shea, et al. (1990). "Sequence requirements for coiled coils: analysis with λ repressor-GCN4 leucine zipper fusions." Science **250**: 1400-1403.

Ingraham, H. A., S. E. Flynn, et al. (1990). "The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions." Cell **61**: 1021-1033.

Ito, H., Y. Kukuda, et al. (1983). "Transformation of intact yeast cells treated with alkali cations." J. Bacteriol. **153**: 163-168.

Kanaan, M. N. and G. A. Marzluf (1993). "The positive-acting sulfur regulatory protein CYS3 of *Neurospora crassa*: nuclear localization, autogenous control, and regions required for transcriptional activation." Mol. Gen. Genet. **239**: 334-344.

Kerr, L. D., L. J. Ransone, et al. (1993). "Association between proto-oncoprotein REL and TATA-binding protein mediates transcriptional activation by NF-KB." Nature **365**: 412-419.

Kessler, S. W. (1981). Use of Protein A-Bearing Staphylococci for the Immunoprecipitation and Isolation of Antigens from Cells. Methods in Enzymology 73. Immunochemical Techniques part B Eds. J. J. Langone and H. VanVunakis. Academic Press. 442-459.

Kissinger, C. R., B. Liu, et al. (1990). "Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions." Cell 63: 579-590.

Koltin, Y., J. Stamberg, et al. (1972). "Genetic structure and evolution of the incompatibility factors in higher fungi." Bacteriological Reviews 36: 156-171.

Kronstad, J. W. and S. A. Leong (1990). "The b mating-type locus of *Ustilago maydis* contains variable and constant regions." Genes Dev. 4: 1384-1359.

Kües, U. (1993). (Personal communication)

Kües, U. and L. A. Casselton (1992). "Fungal mating type genes-regulators of sexual development." Mycol. Res. 96(12): 993-1006.

Kües, U. and L. A. Casselton (1992). "Homeodomains and regulation of sexual development in basidiomycetes." Trends Genet. 8: 154-155.

Kües, U. and L. A. Casselton (1992). Regulation of fungal development by mating type genes. Society for General Microbiology Symposium 50: The Eukaryotic Genome - Organisation and Regulation Eds. P. M. A. Broda, S. G. Oliver and P. F. G. Sims. Society for General Microbiology.

Kües, U. and L. A. Casselton (1993). "The origin of multiple mating types in mushrooms." J. Cell Science :

Kües, U., W. V. J. Richardson, et al. (1992). "The combination of dissimilar alleles of the A α and A β gene complexes, whose proteins contain homeodomain motifs, determines sexual development in the mushroom *Coprinus cinereus*." Genes & Dev. **6**: 568-577.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(259): 680-685.

Lamb, P. and S. L. McKnight (1991). "Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerisation." Trends Biochem. Sci. **16**: 417-422.

Landschultz, W. H., P. F. Johnson, et al. (1988). "The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins." Science **240**: 1759-1764.

Manak, J. R. and M. P. Scott (1993). "Able assistants for homeodomain proteins." Curr. Biol. **3**(5): 318-320.

May, G., L. Le Chevanton, et al. (1991). "Molecular analysis of the *Coprinus cinereus* mating type A factor demonstrated an unexpectedly complex structure." Genet. **128**: 529-538.

Mendel, D. B., P. A. Khavari, et al. (1991). "Characterization of a cofactor that regulates dimerization of a mammalian homeodomain protein." Science **254**: 1762-1767.

Mermoud, N., E. A. O'Neill, et al. (1989). "The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain." Cell **58**: 741-753.

Montague, P. (1987). Comparisons of the molecular structure between mammalian and lower eukaryotic genes. Gene Structure in Eukaryotic Microbes Ed. J. R. Kinghorn. Oxford, IRL Press Limited. 263-285.

Morita, E. H., M. Shirakawa, et al. (1993). "Secondary structure of the oct-3 POU homeodomain as determined by ^1H - ^{15}N NMR spectroscopy." FEBS **321**: 107-110.

Murphy, M. R., M. Shimizu, et al. (1993). "DNA-protein interactions at the *S. cerevisiae* $\alpha 2$ operator *in vivo*." NAR **21**: 3295-3300.

Murre, C., P. S. McCaw, et al. (1989). "A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD and myc proteins." Cell **56**: 777-783.

Mutasa, E. S., A. M. Tymon, et al. (1990). "Molecular organization of an A mating type factor of the basidiomycete fungus *Coprinus cinereus*." Curr. Genet. **18**: 223-229.

Nasmyth, K. and D. Shore (1987). "Transcriptional control in the yeast life cycle." Science **237**: 1162-1170.

Nasmyth, K. A. (1982). "The molecular genetics of yeast mating type." Ann. Rev. Genetics **16**: 439.

O'Shea, E. K., J. D. Klemm, et al. (1991). "X-ray structure of the GCN4 leucine zipper a two-stranded, parallel coiled coil." Science **254**: 539-544.

O'Shea, E. K., R. Rutkowski, et al. (1989). "Evidence that the leucine zipper is a coiled coil." Science **243**: 538-542.

O'Shea, E. K., R. Rutkowski, et al. (1992). "Mechanism of specificity in the Fos-Jun oncoprotein heterodimer." Cell **68**: 699-708.

O'Shea, E. K., W. F. Safford III, et al. (1989). "Preferential heterodimer formation by isolated leucine zipper from fos and jun." Science **245**: 646-648.

Oas, T. G. and S. A. Endow (1994). "Springs and hinges: dynamic coiled coils and discontinuities." TIBS **19**: 51-54.

Oettinger, H. F., R. Pasqualini, et al. (1992). "Recombinant peptides as immunogens: a comparison of protocols for antisera production using the pGEX system." BioTechniques **12**(4): 544-549.

Olson, E. N. and W. H. Klein (1994). "bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out." Genes & Dev. **8**: 1-8.

Otting, G., Y. Q. Qin, et al. (1990). "Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution." EMBO J. **9**: 3085-3092.

Pabo, C. O. and R. T. Sauer (1992). "Transcription factors: structural families and principles of DNA recognition." Ann. Rev. Biochem. **61**: 1053-1095.

Papazian, H. P. (1950). "Physiology of the incompatibility factors in *Schizophyllum commune*." Bot. Gaz. **112**: 143-163.

Pardo, E. H. (1993). (Personal communication)

Phillips, C. L., A. K. Vershon, et al. (1991). "Secondary structure of the homeodomain of yeast $\alpha 2$ repressor determined by NMR spectroscopy." Genes & Dev. **5**: 764-772.

Pillai, R., K. Kythe, et al. (1993). "Use of a yeast expression system for the isolation and analysis of drug-resistant mutants of a mammalian phosphodiesterase." Proc. Natl. Acad. Sci. USA **90**: 11970-11974.

Promega Protocols and Application Guide. (1991). Promega Corporation USA.

Quian, Y. Q., M. Billeter, et al. (1989). "The structure of the *antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors." Cell **59**: 573-580.

Rambaldi, I., E. Nagy Kovacs, et al. (1994). "A proline-rich transcriptional activation domain in murine HOXD-4 (HOX-4.2)." Nucleic Acids Res. **22**(3): 376-382.

Rao, P. S. and D. J. Niederpruem (1969). "Carbohydrate metabolism during morphogenesis in *Coprinus lagopus*." J. Bacteriol. **100**: 1222-1228.

Raper, J. R. (1966). Genetics of sexuality in higher fungi. New York, The Ronald Press.

Raper, J. R., M. G. Baxter, et al. (1960). "The genetic structure of the incompatibility factors of *Schizophyllum commune*: The A factor." Proc. Nat. Acad. Sci. USA **46**: 883-842.

Raper, J. R., M. G. Baxter, et al. (1958). "The genetic structure of the incompatibility factors in *Schizophyllum commune*." Proc. Nat. Acad. Sci. USA **44**: 889-900.

Richardson, W. J. R. (1992). The molecular analysis of the A mating type factor of *Coprinus cinereus*. University of London.

Rogers, S., R. Wells, et al. (1986). "Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis." Science **234**: 364-369.

Rost, B., R. Schneider, et al. (1993). "Progress in protein structure prediction." TIBS **18**: 120-123.

Saenger, W. (1984). Principles of nucleic acid structure. Heidelberg, Springer Verlag.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular Cloning. A laboratory manual. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain terminating inhibitors." Proc. Natl. Acad. Sci. USA **74**: 5463-5467.

Sauer, F. and H. Jäckle (1993). "Dimerization and the control of transcription by *krüppel*." Nature **364**: 454-457.

Scherer, S. and R. W. Davis (1979). "Replacement of chromosome segments with altered DNA sequences constructed *in vitro*." Proc. Natl. Acad. Sci. USA **76**(10): 4951-4955.

Schulz, B., F. Banuett, et al. (1990). "The b alleles of *U. maydis*, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif." Cell **60**: 295-306.

Scott, M. P., J. W. Tamkun, et al. (1989). "The structure and function of the homeodomain." Biochim. Biophys. Acta **989**: 25-48.

Sessa, G., G. Morelli, et al. (1993). "The Athb-1 and -2 HD-ZIP domains homodimerize forming complexes of different DNA binding specificities." EMBO J. **12**: 3507-3517.

Shariari, H. and L. A. Casselton (1974). "Suppression of methionine mutants in *Coprinus*. I. Complementation and allele specificity of suppressor gene action." Mol. Gen. Dev. **134**: 85-92.

Shirakata, M., F. K. Friedman, et al. (1993). "Dimerization specificity of myogenin helix-loop-helix DNA-binding factors directed by nonconserved hydrophilic residues." Genes & Dev. **7**: 2456-2470.

Simon, H. (1974). Messung von radioaktiven und stabilen Isotopen. Berlin, Heidelberg, New York, Springer Verlag.

Smith, D. B. and K. S. Johnson (1988). "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-S-transferase." Gene **67**: 31-40.

Smith, D. L. and A. D. Johnson (1992). "A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an $\alpha 2$ dimer." Cell **68**: 133-142.

Sorensen, P. H. B., L. Lessnick, et al. (1994). "A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS-family transcription factor, ERG." Nature Genetics **6**: 146-151.

Southern, E. M. (1975). "Detection of specific DNA sequences among DNA fragments separated by gel electrophoresis." J. Mol. Biol. **98**: 503-517.

Specht, C. A., M. M. Stankis, et al. (1992). "Functional analysis of the $A\alpha$ homeotic mating-type genes of *Schizophyllum commune*." PNAS USA **81**: 7174-7178.

Stankis, M. M., C. A. Specht, et al. (1992). "The $A\alpha$ mating type locus of *Schizophyllum commune* encodes two dissimilar, multiallelic, homeodomain proteins." PNAS USA **89**: 7169-7173.

Stoscheck, C. M. and J. King L.E. (1986). "Functional and structural characteristics of EGF and its receptors and their relationship to transforming proteins." J. Cellul. Biochem. **31**: 135-152.

Struhl, G., K. Struhl, et al. (1989). "The gradient morphogen *bicoid* is a concentration-dependent transcriptional activator." Cell **57**: 1259-1273.

Studier, F. W. and B. A. Moffatt (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." J. Mol. Biol. **189**: 113-130.

Studier, F. W., A. H. Rosenberg, et al. (1990). "Use of T7 RNA polymerase to direct expression of cloned genes." Meth. Enzymol. **185**: 60-89.

Sun, X. H. and D. Baltimore (1991). "An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers." Cell **64**: 459-470.

Swiezinsky, K. M. and P. R. Day (1960). "Migration of nuclei in *Coprinus lagopus*." Genet. Res. Camb. **1**: 129-139.

Swiezynski, K. M. and P. R. Day (1960). "Heterokaryon formation in *Coprinus lagopus*." Genet. Res. Camb. **1**: 114-128.

Tabor, S. and C. C. Richardson (1985). "A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes." Proc. Natl. Acad. Sci. USA **82**: 1074-1078.

Tanaka, M. and W. Herr (1990). "Differential transcriptional activation by Oct-1 and Oct-2: Interdependent activation domains induce Oct-2 phosphorylation." Cell **60**: 375-386.

Tapscott, S. J., R. L. Davis, et al. (1988). "MyoD1: A nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts." Science **242**: 405.

Taylor, W. R. (1986). "The classification of amino acid conservation." J. theor. Biol. **119**: 205-218.

Tomei, L., R. Cortese, et al. (1992). "A POU-A related region dictates DNA binding specificity of LFB1/HNF1 by orienting the two XL-homeodomains in the dimer." EMBO J. **11**: 4119-4129.

Towbin, H., T. Staehlin, et al. (1979). "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications." Proc. Natl. Acad. Sci. USA **76**: 4350-4354.

Tymon, A. M., U. Kües, et al. (1992). "A fungal mating type protein that regulates sexual and asexual development contains a POU-related domain." EMBO J. **11**: 1805-1813.

Verrijzer, C. P., M. J. Alkema, et al. (1992). "The DNA binding specificity of the bipartite POU domain and its subdomain." EMBO J. **11**: 4993-5003.

Verrijzer, C. P. and P. C. van der Vliet (1993). "POU domain transcription factors." Biochim. Biophys. Acta **1173**: 1-21.

Vinson, C. R., T. Hai, et al. (1993). "Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design." Genes & Dev. **7**: 1047-1058.

Vogt, A. B. (1992). (Personal communication)

Voronova, A. and D. Baltimore (1990). "Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains." Proc. Natl. Acad. Sci. **87**: 4722-4726.

Watt, F. and P. Molloy (1993). "Specific cleavage of transcription factors by the thiol protease, m-calpain." Nuc. Acids. Res. **21**(22): 5092-5100.

Whitehouse, H. L. K. (1949). "Multiple-allelomorph heterothallism in the fungi." New Phytol **48**: 212-244.

Wolberger, C., A. K. Vershon, et al. (1991). "Crystal structure of a MAT α 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions." Cell **67**: 517-528.

Xue, D., Y. Tu, et al. (1993). "Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3." Science **261**: 1324-1328.

Yashar, B. M. and P. J. Pukkila (1985). "Changes in the polyadenylated RNA sequences associated with fruiting body morphogenesis in *Coprinus cinereus*." Trans. Br. Mycol. Soc. **84**: 215-226.

Yasuda, H., A. Mizuno, et al. (1994). "ATBF1, a multiple-homeodomain zinc finger protein, selectively down-regulates AT-rich elements of the human α -fetoprotein gene." Mol. Cell. Biol. **14**(2): 1395-1401.

Yee, A. R. and J. W. Kronstad (1993). "Construction of chimeric alleles with altered specificity at the b incompatibility locus of *Ustilago maydis*." PNAS USA **90**: 664-668.

Yuan, Y. O., H. L. Stroke, et al. (1993). "Coupling of cell identity to signal response in yeast: interaction between the α 1 and STE12 proteins." Genes & Dev. **7**: 1584-1597.

Zappavigna, V., A. Renucci, et al. (1991). "*HOX4* genes encode transcription factors with potential auto- and cross-regulatory capacities." EMBO J. **10**(13): 4177-4187.

Appendix I. Alignment of nucleotide sequences of *b* gene-pairs
of A6 (top row) and A42 (bottom row)

CTCGAGTTAC	GAGTTATTAC	TCTCTACNAC	NTTTCCTTT	40
.....				0
CCAACGCcAC	TCCTGTTTGA	CNGGGAAAGA	AAGGCTGTAG	80
.....				0
TGACTAGCCA	CAAGCAATAC	CGTGACCCAG	AAACCCTCTA	120
.....				0
AACTTATCCC	TTaAAgTCGT	CATCTTCTCT	CCGTCACTTC	160
.....				0
GCACTCGGAT	TGCCTTACAA	TACGTGGAGC	GCATCTCTCG	200
.....				0
AAGTCCCTgA	CGTCGAATAT	GATGACTTgC	TGGGTACCTC	240
.....				0
ACGTACcCTT	TAcTGATTGT	TCTTTAGCAC	CAGgcaccca	280
.....				0
ancaCTGTGC	TTCGTGGATG	CCTGTGCCTA	GTAAATTgAT	320
.....AAGC	TTCGAAGGATG	CCTGTGCCTA	GTATTTCGCC	34
gCACTGCATA	TGGAGGGAAG	CTCTAAGAAAC	TCGGCnnnnng	360
GTACTAGATA	CGGAGGGA..	CATCTAGAAA	CCTGGCGACG	72
aaaaaccagta	ctccgaacct	gtaatttccaa	gcagg.aagca	399
AAAACCCAGTA	CTCCG.ACTT	GTAATTCCAA	ACAGTAAGCA	111
acagccagttt	cgTCcAGTca	TCCAGCCTCA	TCCAGCCTCA	439
ACGGCCGAT.	..GTcATCCa	TCCAGCCTCA	TCCAGCCTCA	148
AA.....	GG	CCTAggTTCT	CTCAgCAGGC	463
AAAGGGCCTA	TACTGCCCAg	ACCTGGATCT	CTCAACAAGC	188
GTGAAGTGAG	ATACA AAAAG	CATCATAGTT	TCCAGCCTCA	503
GC GAAGTGAG	GTAC. AAAAG	CATCATAGTT	TCCAGCCTCA	227
GGTA TGTT TG	ACATAGTGTA	ACAATCTAAA	CTACTA ACTT	543
GGTACGTT CG	AC. . AGTG TG	ACAATCTAAA	CTACTA ACTT	265
CTAATgggtac	actagaggCAa	taggat aatc	AAagacacaa	583
CTAATGGGTAC	A ACTA AACAG	CTAATAAATC	AAGCAAGGAT	305
tga c gat agt	atgctggatg	gcta. gaata	caattggcga g	622
GGT CGATAGT	ATGCTGGATG	GCTAG GAATA	CAATTGCGTC	345
Gagttttgggtt	tGC. . Gaaac	aaAtgggaata	.aaAtgaagaa	659
GAGTTTAA GTT	TGCAGA AAAC	AAATC GAATA	GAAATAAAGT	385
tggaaaGCGA	AGATCGaAAG	CATAGCGACA	GCAAAGCATG	699
TCGAAAGCGA	AGATCGAAAG	CATAACGACA	GCAAAGCATG	425

GAGGGGTATCa	acGACGT . . g	GGTATATAGC	AGCGTAGGTC	737
GAGGGGTATT	ACAACGTGGG	GGTATATAGT	AGCGCAGGTC	465
CAGCAAAGGA	TCAAGGGGGCG	CaGATTTCGG	CGCGAATGGC	777
TGCCCAAAGGA	TCAAGGGGGCA	CAAAATTTCGG	CGCGAATGGC	505
TTGGGCTCTC	GCTTCAACTT	AGCGGAGGTC	GC TGGCCTTG	817
TTGGGCTCTT	GCTTCGACCT	AGCGAAGGTC	GT TGGCCTTG	545
GCTTCACCTG	CCTGTTCAGA	CAACGCCTTTG	GATGTCGacA	857
GCTTCGCCCTG	CCTGCTCAGA	CAACGATTCGG	GATGTCAAAC	585
ACAACCTGATG	CGAAAGGAAC	TGAATCGACG	CgGCACA CGGC	897
ATGCCCTGCTG	CAGGAGGAAC	TGAATCGACG	AGGTCA TAGC	625
AATaCCCGTC	GAGCACCTGGC	CAAAACAAGCC	AGAGAAGTcg	937
AAAATTGTG	GAGGGTTGGC	TGAACA CTGT	AGAGAAGTCG	665
agatcaggat	ga gagatgtt	ggaccatage	gat tgggtcat	977
AGATCAGGAT	GGGAGATGTT	GGACCATAGC	GACCGGGTCAG	705
ggaggaaagc	agagGTCtga	gtgtccgcga	aaggt tgcga	1017
GGAGGAAAGC	AGAAGTCCGA	GTGTCCGGGA	AAGGTAGTGA	745
tataCaATCT	aatcttggga	caagaggttga	atgcaaatAA	1057
CATCCAGTCT	AATCCGGA GA	ACAGA		770
ATGTTGGCTG	AGCTTgGGAG	TGTATTGGTC	GGTTCTGAGT	1097
. GATG	AGCTTGGGA A	CGTATTGGTC	GATTCT T	800
GAGTGGTATC	AGTTAAACGT	gGTTCACTGC	AGA TCACA GG	1137
GAGTGGAAAT	AGTCGACGTG	GTTC CACTGC	GTG TCACA AT	840
ACTTACGAAG	CAgTGGAGGA	AAATCA T TGG	CCATCCACGA	1177
ACTTACGAAG	GGGTGGAGGA	AAATCAC CTGG	CCATCCACGA	880
CAGCTTTTGA	TCGaAGGGTTG	GTA Cg TCAGG	GTCAAC T TGG	1217
CAACGTA TGA	TCGAGGG CCG	TAGCT TCAGG	GTCAAC G TGG	920
GAGTGGTtgt	cga cccggtac	ttca agtttg	tcgata agat	1257
GAGTGGGTGA	CAGCCGGGTAC	TTCGAGTGGG	TCGATAA TAT	960
TAggCTCAGC	TCGTggTgCT	. AcAGTA gTA	GTCAGT C atG	1296
TGAGTTCAGC	TGT TGGGTGCT	AA TAGTGA T G	GT TGGTGA TA	1000
Gt . . CACaCa	ATAGACAAGC	AAAGATGTCTT	AGTGCTCGGC	1334
CTCGCGAATA	ATAAACAAGC	ACGATGTCTT	AGTACCGGACA	1040
A . . . GAGTGG	AgGGAGGGAG	ATCAGGGATG	AGA gggAAAT	1371
ACAGGAGTGG	AGGGAGGGAG	GTCAGGGATG	AGG GGGAAAT	1080
TgAACAGTTG	AATATCGAGA	AGG GTGTGTGG	AGAGCggGAG	1411
TGAACAGTTG	AATATCGAGA	AGAGTGTGTGG	AAAGAGGAAG	1120

C T C G A G C C T G	A G C C G T G A G T	G A A T G T C T C A	A G T A A C G G T C	1451
G T C G G T C C T G	A C G C C C G A G G	G A A T G T C C C A	A G T G A C C G T C	1160
A C C T C A G G A G	A G G A G A G A A C	A T G T T C A C G G	A A C C A T T G G T	1491
A C T T C A G G A G	A A G A G A G A A C	A T G T T C A C G G	A A C C A C T G T T	1200
C C C A G T C T T G	C g C A T T C G C A	G C G G G G A A A G	G A T c C G A A A C	1531
C C C A A T G T T C	G G T G T T C G T A	G C G G G G A A A G	G A T C T G A A A C	1240
G G C C T G T g g c	g A C C T G T . . .	G G T G C T T T G T	A G C A G G A T A C	1568
A A C C T G T G G A	C G T A T C T C T G	G A C G C T T C G T	A G C A G G A T A C	1280
T G G A A G C C A T	C G G A T A A G C A	T C G t T T C C G C	T G G c C C C G T G C	1608
T G G A A G C C A T	C G G A T A A A C G	C C G T T T C C G C	T G C C C C G T A G	1320
A T G T C G A C G A	c C g g G T C g g G	G t G	. . g G A G A C G A	1639
A T C T C G C C G A	G T C A G T C G G G	G T G A G C T T A G	G A G G T A G C G A	1360
C G G T T G A G A T	C A G G C A A T G g	G G G A A g A T T C	T A A A A G C T G C	1679
C G G T T G T G G A	C A T G G A G T G G	C G G A G G T T T C	C A G G A G T T G A	1400
T C T T G A A T C G	A G g A G T C T g G	G G A G G G C A G T	C c T T T C A G G G	1719
T C T T G A G T C G	A A G G G T A A G G	G G A A G G C A G T	C C T T T T A C A G	1440
G A C T G A G A T T	A C G A T G C A T G	g a C A C A T C T C	T G C A G C G A A G	1759
G A C T G A G T T T	G C A A A A C A C T	G A T T C G G A T C	T G C A G A T C C A	1480
A T C A G A G A A a	g . . C C C A G A C	A G G A A G A G G G	T A C A T A C C G G	1797
T T C A G A A C C T	G G T T T T T A A A	G G G A G T G T G A	T A C A T A C C T G	1520
G C T T C T T T T C	C A A C c g G g a g	c g a g g T A T C T	G G A T G A G T G T	1837
G C T T C C T T G C	A G A C T G G A G A	T G A C A C G G T T	G G A T G G G T T T	1560
G C a A T G G T C a	A C G A C G C a t g	c g a c t c a g g a	t c g G t g a c g t	1877
C C A G T G A A C A	A C G A C G T A T A	C G A C C A A C G A	A T G G C A A T T T	1600
g T G g a t g t t c	a a c t c g a t g a	C a t g t g a c g a	t g G g g G g g t t	1917
A T G T	G A A T C G A T C G	C G A A C G A T G A	G G G C G G A G C T	1634
a a a a g c t g g G	a a g g t C a A G C	A G G C a A G C G T	T G A G G A T G G G	1957
A A A A G T T G A G	A A G C C C A T G C	A G G T G A G C G T	T G T G G A T G A G	1674
C C T G A G C G T G	A T G g g c g c g T	. . . T T C C T C A	T G C G A G A a T C	1994
C C T G C G C G T G	G T G A G C G C G T	T G A T T C A T A A	T T C C G A C A G T	1714
T A t G T C T T c c	t g a c t g t C G c	T A G C C T T C T C	T T T C T T C A G A	2034
G C T T G C T T C T	C G C T T G C G C C	T G G C C T C G T C	G T T C T T C A G C	1754
G c C T T c C G C a	T G G A C G G A G T	C A T T G C C T T G	A G A A G A G A A T	2074
T G C T T C T T C A	G G G A C G G T G T	C A T A T C C T T G	A G T A G G G T G T	1794
C C A A C T T G T C	A G C G A G T T T G	C T C T G C G A G A	A C T T C T G G t g	2114
C C A A C T T G A C	G G C G A G T T T G	C T T G G G G A G A	A C T T T T C C T G	1834

GTA GAAG GTC GTA AAGCGAT	GTTGGCA TGT C CTGGCG GTGGG	CAGGA ATTGG AAGGG ATTCG	GGCGAA T TGA TGGCGAAC TGA	2154 1874
CTGTC GATAT AGGTC AATGT	CGG CCTGCAA CATCAGGCAA	GAGATCTGAT TGGGTCTCACC	ACCAACGGGA ACAGCGGGGA	2194 1914
CCAGGTCTCA TCGAT TGC CG	GTTGAAA AAG ACC GG TGAAA	ATAGAAGCAG ATC GAAGCTG	GTTgGAT TAT GACGAATGAT	2234 1954
CtTCGCCTCAT GTCAAACGCA C	TTA TTTGGGGA TTGT TTTGGGA	AATGCTTGCG AATGCTTGCG	GCGCAGATGA GCGGACCTGA	2274 1994
TGGAAGCCAT TGGAAGCCAT	TACGCCTGCG TTCGCCTGCG	AGGA TCAATA AGGG TCAATA	AACCAAGCAA AACCAAGGCAA	2314 2034
CGATGTTCTT CGATA TTT TTT	CACGGAGTTC CCGGGAAGTG	GAGTTCTTGGT CGGGTTTGGC	TTgGGATTTTC GCGGAATTTG	2354 2074
TGAACGA ACT GGTC CGA ACT	TGCGgGCTGG TGGCGA CTGG	GGTAGGGGTT GGTACGGGTT	GTGAACAATTA GTACCAAGTTG	2394 2114
TCCTTGAGCC TCCTTGAGCC	aCCAT CCGGT AGCAC CCGGT	TGGCTCTTAGG CGGTTCGACA	CATGACGGAG CAGGCGGGAG	2434 2154
. TTG TAGGC GA TTT	ATTCATCTAT CGGTTCCTGC	AaCATTTTGGGA CAATGCGGGGA	AGAAACCCATC AGAAACATAG	2467 2194
TGTAAaTTGa AGTAA C GTCA	 GATAGA AGTCAATGTC	AaTACGGACc CCTCAGGATT	GTtATACTGG CTTATCTCTGG	2503 2234
TCACCGACAT GAACGGTCA A	tCAGTTGACa GACC TGA GCA	 GGAG CTTGTC TGA A	aACAGgCCAA AGCATGGCCA	2537 2274
CGAGATTTCaA TGACCCTCCAC	TGAGATTTC C TGTGATCTTC	CCGGATGGAG TTTGATTGGG	GCTGTGTCTG GT TGC GTTTG	2577 2314
TGCGTTTCAGC CGTT TTTGAGT	TCGAGGCTAT AAAGG GCTGT	TGCTCGAAaG TCTCAGAAAAG	gCTGGATgGT TCCGA AAGGC	2617 2354
GgCGCCACA A GATT C GGAG A	CaTCGtCCAC CAGAA TAGAG	TTCGTCCAA T CTCA TCGAGG	GT TGCATGAG GTGGCGTGAC	2657 2394
TCAGAGATTC TCAGAAAT TCC	CTCAGCGCaC ATCAGCA CAT	ACAAACAaGAG TCAGCGAAAAC	AGTCGAAATTT CGTCGAACTTT	2697 2434
CAGCTTAGGCG GGTCC CAGGCG	CTGAGGAAAg GAGAGGAAAGG	CATCAAAGGA CAGAGTAAGC	GGTCGATTCC AACCGATTCC	2737 2474
tgaCagAGgA TGTCA CAA CA	GGCAGAA GAA ATCCAGACAG	atCtTTGGGA ATCCGCGGCGC	AGGCtgcTtAA AGCGTGT TGA	2777 2514
TGgTAGTCgT TTGCAGGTGG	CACAA tGAAA AATGT TATCG	TGgctGTCTT TCACGACTAT	CTTCCGACAA CGCATG CCGA	2817 2554

AACATGCGTA	GCAACAGGGA	TTCGGGTGAG	ATGCAGGAGA	2857
TGTTTCGCGCG	GATACC GGGGA	TTCGTAGGGG	ATAGCCGCGA	2594
.. .. .	 GA	TCGG AATCTC	GATGAGACGA	2879
TAGGT TGTGA	AGGT TGAAGA	TCAGGCGTCC	AATGATCCCA	2634
GGCGAACTTGA	ACTGCAACA A	TGCGAGAAAG	TGGTTATGTC	2919
GCTTTCCCGA	TGTG GTCGGA	GTTTCA AAGG	CGCAGGTGA A	2674
GGAGTGCGAA	AGACTTGCgG	ACAGGTG aCA	GCA T TCA TTG	2959
CGTATG GTAC	CCCAAAGAGG	TCAAGTTTAGT	CCA A A CACA	2714
TCCTTCA AAG	ATCTCCAAAC	ATTCTCATGC	AATCA	2994
GCAAAACATAG	TCGACAGTCC	ACTGTAAATAT	TAACA TGGGT	2754
.. .. .		 ATGGAT	GTTTCGAGTCC	3010
TGTGT TGT TG	TTACAGTGGT	AATC AAGGAC	ATTTGAGCAA	2794
gGTC AATgTC	TGg			3023
GGGCCTGGA A	TGG ACTTGAA	CAATATTGTT	CTAAAGAA GC	2834
.. .. . A	GGCGGT TGA G	CTATAAGTTT	TACCTCAAGA	3054
GAA C GATGC A	TGGCCTTCAGA	ATCGATGCTG	GTA CTCA CGT	2874
GGAGTG GTACC	AGGTAC			3070
TGAAGGT TGT	AGGTACCACT	ACATAGTGT C	AATCTCGAGG	2914
.. .. . GATG	gTCTCTCTAT	ATT	TTT TGg GTAC	3097
GAT TGA GTTC	GGCGCTCTAC	ATTGAGACC A	TGG TGGATAT	2954
TCATAAT TGA	GATAGGCGGTT	AC A	AGTGCGGCCCa	3130
GTATCGTGTCT	TGAGTGA GTA	ACACGGGA A A	AGTCAAGCA CA	2994
TATTCTAC . . .			 aTt	3140
CACGC GCA GA	CATGT TATTT	CGACGGAAGG	GCAGTGA ACT	3034
CCTTCA TACA	C TAG	GCTGAATCGA	AACTCCGCA T	3174
TCTTCTGTACT	CGCGAGCAAG	GGCAGATTAG	AACTCCGCA T	3074
TGTCGt CCAA	ATA CACGGTC	GATAGCAATC	TCCGCGGCGG	3214
TGTCCTTCAA	ATAAACAAGCG	GTGTCCTCTGC	ACGGAGGCTGA	3114
ATTCGCTGTG	TTGGCGGCTTT	CAAAATAC . .	 ATATTTC	3248
CGGTCGGATA	TATGAGGCTG	CATC ATACA G	CCCA ATGTTG	3154
CCC CATCTC	TTACGTGAGC	 CCA TA	TGCAGCAAGG	3284
CTGCACGATT	TTCTGCTTAC	TCTG CACAAA A	TGATACGAGT	3194
ACC AAACGGT	GACCTATG GGG	GAGg ATAGTC	AGAAATGTCTG	3324
CGA AAACCCC	GAACTCTG GGG	GCGGGCAGCG	AG . GAGCTAG	3233
CAGTCA TGGG	CACGGTTAAG	ACGAAGTGTCT	CTGCGGCCAG	3364
AGGCTCTGGA	CTCGTTTGAG	TCAGAGTGTG	C	3264

AGTCTGGCCTGCCAA	AGACACGCGC AACCAAGCGT	ACGCAACGCG CTGTTCAGCG	TTCTCTTTCT CAGGGGGTTA	3404 3299
AAATCACGTT CAGAGTTTGT	AAATTCCTTGT AATTCCTCTT	GCCAACCTGT GCAGCCCGCA	CATCTTCTGT AATCTTTTGT	3444 3339
cgaCCCGTCG CGACTTCCCTGT	ACTTCAATGGA CCTTCACTCA	GATcATTTCAG GATCGTTCAGG	CAACTGGGGGC TCACTCGAGGC	3484 3379
TTCTGAGCAGC TTCTGGGGCCTC	CCCAAGTGGGC CCCAAGCGAA	AAACTTTACCG CAACTTGGTAC	TCAAGATCga TGAAAGTTGGA	3524 3419
CAGTGCCGTC CAAAGTGGTC	CAGCAGCAAA CAAGAAACAAC	GACAAAGCTCG GCACCTCTTCA	GCTCAAGACT TTTAACCAAC	3564 3459
TATCAAGAGA TATCAACCGA	TGTGCAATACG TGTGCGGAGG	ATACGTCGAA ATACCCTCGAG	ATAAGTCAGC TTGAGCAAGC	3604 3499
AGAAATCAGG AGCAAGGAAT	CCCCATCGAT GGAAATCGAC	TCCAGGGCTG CCGATTTCTG	TCGACCAATGC TCGAATAACAT	3644 3539
TCGCAACTTT GCAGACATTTG	TGCGGAGGGGT GTGGGGGGCT	GGGTGTCTCA GCGTCCAAACA	GGATCTCGCA ACTAATCTCA	3684 3579
GACATCAACT CGAATTCACT	CCGCATTTTAT CGGAATAAAT	GGATTGTGAG GGAACGCGAG	AATCGCTGTG CAGCAAAACC	3724 3619
TGCAAGGCCCA GGCAA...CAA	ATCTCGTTCTG ATGGAAATTCA	GAAgGGAAGS AACGGGACACS	GTGCGCGTTT GTTCAGTCC	3764 3657
CAACCAATGA GTAAACAATCTT	AGCTATCGAG GTATTATCCT	TCGGGCAGAC TTGCCGGCCC	CTTGCGGATA GCCGCGCATTA	3804 3697
CTGCCGCAAT GTCCTGACTC	AGAGAAAGTcg AGAGAAAGTCG	GtGCTTTCAT GTGGCTTTCAG	GAGAAAGTACT GAGAAAGTACT	3844 3737
ATGAATACAA ATGAGGTACAA	TACCCTCAGGA TACGTCAGGC	TCGGCTCTTGG TCTGGCCCTGG	ACCGCGAATG ATCGCGAATG	3884 3777
AGTTGCGGCC AGTGGGCCGCC	CAATCCATAC CAGACAATAC	TTCTGAACGCG TTCTGTGTGCG	TCAAATTCGAA GCAAATTCGAA	3924 3817
GTTTGGGTAA GTTTGGGTAA	CTCTTAAttcA CTGTTTTTTCT	GGCCTCCACG CGCCTTGACC	TGTTGACTCg TTGATAATCTG	3964 3857
TTGTGCCC GG TTCCTGTCTGTG	CCATTTTAGTT ACATTCTAGTT	CGAAAAACCAC TGAGAAACCAT	agaCCTGGCa AGAACTGGAA	4004 3897
AGCCAAGGGT AGACAAAGGT	CGGAATCCAC GGGAATCCAT	TTTGAAAGAC TTCGAAAGAC	TGCCCATGGA TCCCATGGA	4044 3937
CCaATTAGcc TCACTTTTGCC	GCGCACATAT GCCCAGCTTT	GACtAgaata GCTTGGAATA	CGAGGAAAGGC TGAGGAAAGA	4084 3977

CA CGt CC CGG	GCTTCACACt	ACCAGGCAAA	TCCGAGGCCA	4124
CAGCTGTCCG	GATTGACATT	GCCGGACAAA	GTGGAAGTTT	4017
GC GAGTCAGC	TTGGCCGACG	CCTCCCTGTAA	GCCGGTA t CA	4164
ATGGCTCAGC	CCTGCCAACAA	CCTCCCTCGTAA	ACAATCGTTA	4057
ATGGTGTCAC	TACATCTTCA	ATCTGACC CC	GA . . . CAGAG	4201
TTTGAC TGT	CCTTCCTT .	GTCCTGACCAG	GATAC CCGAG	4095
GTACAAGTCC	ATCCTCAGGT	TTCATTATCG	CCGAGGAACA	4241
GTGCA TGC GA	ACCACATCTG	TTCCTCCCTG	ATTGGCCTAG	4135
AGCCAACATG	TGCTGGCGGC	CGTGCC CCAA	CTCATGATGA	4281
TGTCACTCTGA	CGAAGGTGGG	GGGCAT CCAA	CTCAGGATAG	4175
CATTCTTGCC	gGAGAA CACG	CCTTTCCCCAC	TATCTTCCCG	4321
AGCC C CAGGC	GGCCGG CACG	CATATCCCCAC	TGTCTACAGA	4215
AAAGGCAGAT	CGGACAAGAA	acGTCTT acC	TTCAAGCATTA	4361
AAAGATGGAC	AGAAATGAGAA	ACTTCTC ACT	TGCAAAACGTA	4255
GAGATTTCAA	TTTCCCTGCT	GtGGTATG GG	ATCGTCTGACT	4401
GGAATTTCAA	TTTCCCTGCT	GTGATT TTGGG	ATCGTTGCTC	4295
TGA AAACGCT	GTCCGGGACAC	TCTCTTTGGAC	TCCCCACCGTCT	4441
GTC AA . . . CG	ACC C GACCAA	AA TCTC GGA	TCCCCACAAATG	4332
CATGAAGTGT	GTGATGCACT	CGTCTCCGGA	GGTATTCTGTG	4481
GATCAGGTAT	GTCAAAACATT	TGCGTCTTGG	GGCATTCGTG	4372
CC CAGC C CCG	AGCCGATAGAA	GCC AACCAAC	CAAGCGCAGC	4521
CGCAATCGAA	GGCCCAAGAA	GCC CCGGGTC	AACGTG CCGC	4412
TTGGTATTCa	TGA CAAATTCT	CTGGT CAGAT	AAAAGTGGAG	4561
TTGGTATGCA	TGT CAGTTCA	CTGGACATAT	CATTGTAGAG	4452
GCTTCGGGCCT	TAAATCGCCTT	TCGTAA AACCC	TCTCTTTCTCT	4601
GCTTCCTG CAG	CGATAGCCTC	CTGTGAACCA	AGATTACCTC	4492
TAGGA A . . . G	GGCTTCGTCT	GGCTCAATCG	TCAGGGGATG	4638
TCGAG AAGTG	TGT TTGGCTCT	CGAAGACCAT	CGCGAGGATG	4532
AGCGGAACAG	CGCCGGGGA CC	CCCGA ACTgG	AatTTCTGCA	4678
C GCTTCACCG	CGCCGGGGAAC	CACAAGTCGG	TATCC CAGCA	4572
CCCAGAGGCTt	CGAA CCA . Ca	TCCACTATCTG	GCCTTG C GCT	4717
CCCAAGGCTT	CCCAATCACCA	ACAACAACCA	CGCTTACATA	4612
CACCCACTCC	GTCTCCCTCC	CTCACTTGGGA	ggGCAGGGCA	4757
CAACCACTCC	ATCTCCCTCT	CTAACTT CGA	GGTTAGAGAA	4652
ACGAGCCACG	CCCTCCCTTA	AGCTCTTAATA	CATTCAGGtg	4797
ACGCACGACG	. CCTCCCTTTG	AGGTCTCAAA	CATCTAGATG	4691

TTCCGCGAAT TTTGTGCGAAC	TTGACTGACC TTAAGTGACC	TCGAACACACC TCGACACACCA	TGACGCCCTCG GGATGCTTCG	4837 4731
CTATCAATGG CCTTCAATGT	GGCCCGACCT CGCCCGCCTC	GTAATTTCGG AGCATTTCGG	GGCAAGAACG CCTATGGTTG	4877 4771
AGCGAACCGC AGGAGCATAT	GGAGAAACGAC CGGGGATGCT	TTCAACGCGCG TATGCTGTAA	AGAAACCCGCA AGAAATCAGTT	4917 4811
GAGAAAGCTG GGACAAGGCG	CGTTGCGA AGCTGGGGCAG	GTTATCAAGAC CCATCTACCT	CCCTAGTGCA TCAAAGCCCT	4957 4851
TCTGATAACA TCTAACAACA	GCAACTTCGC GCAGGCCTTG	TTCCGAGCGC TTCTGGGGCA	GCCCTTTAGCT CCCTTTTGGC	4997 4891
TAAAGTTTTC TTA.AATCCC	CGCGTATCGA TATTCATCGA	TTGACATAGA CTGACATA..	TAACTCCGTCA	5037 4919
CGCGCGCGCT CCCTAGTAGG	ACCTGCGACT GGTCTCCATT	CGCAATGTGC CTCACCTGTCC	ATCGTGGCAG AGTGATGTGC	5077 4959
ATGAA.GGGA CTGAATCGAA	CGAGTCGGCC AGCCTCCTGT	AATGAAATA TTCGCCAGCA	TATGTGGTCG CATGCTCTCG	5116 4999
TTTACAAAGGA GTT.....GCA	CAGTATCCAA GATTGCTCAA	GGACGGGCACC GGCTGGGCAC.	AGTCAGTTCG	5156 5024
CACTGTATTA CCCTCTAGTC	CGTTTGTTCT TGTCCAACCG	TCGTCTATCA ATGACAAACA	CGCTTTCT.C CTCTTCATGA	5195 5064
GTTAATAATT GTCATAAGCT	ttcCTAACTG TTCGTTAACT	ACCTGGCTCT TTAGGTTTCT	TTAGTTTGA TTGGTGTCG.	5235 5103
ACTTTGGTGA ..TCCGCTGA	ATTTGATAC ATTTGATAC	GGACGTCCAC GGA.....	CGCATTAGAG TCCATTGGAG	5275 5134
TGTATTCTCTA TGTATTCTCTA	CGCCAGGCTG AGTCGGG.ATT	CCAGTGTATC AAGCTGTATC	TTTACTGAGT ATTA.....	5315 5167
ATACCTAGTtT	GCAACATTC GCTTCACTCC	TACGGTCAAT TTATGTGGAC	GCTTACTTCC CCGATCGCTA	5355 5198
AAACAAACCC ATGCAATACA	TTTTCCTCGT TTCGTCAGGG	AGGATACTTT TGAATAATGT..	CAATACGATG ...TACCCCG	5395 5233
CCGTCAATCCG ACTTTTCAGG	TCtCAGCTCT TGTGTCGATT	CTCTCCAATGg TGCGAAGTTG	AAAAAG..CTA AAAAAGCACAG	5433 5273
CTGTACGCAT AGCTGTGTAT	CTGCCCTAATC TTTGATGGTT	AAAGTGtGCA GAACGTTGA	TGTTACAT.. CGTTGCACTC	5471 5313
AGATACGTGT TGGGACGTTT	ACTCCATAAA GGATCATAGA	CTTTGTACGT CTATTA..T	ATAGTATATT TGGCGAAACT	5511 5351

T	C	C	T	G	A	A	A	T	C	T	C	A	G	G	T	G	A	A	A	T	C	A	G	G	C	A	A	A	C	C	A	G	A	C	5551							
C	C	A	T	G	T	C	A	C	T	A	C	C	A	A	G	T	G					C	T	T	G	A	C	T	A	T	C	C	A	T	A	C	G	T	C	A	G	5391

C	T	T	T	C	A	A	G	T	T	T	C	A	G	G	A	A	G	C	T	T	.5572
G	C	T	A	G	C	A	G	T	A	T	G	A	G	G	T	T	G	C	A	T	T5413

Appendix II

Partial DNA sequences of hybrid cDNA clones

As mentioned in chapter 3, two of the cDNA clones that were obtained by screening the λ ZAP library contained hybrid inserts. DNA sequencing of these clones revealed that the pb1-3cDNA clone contained some 0.6 kb of non mating type cDNA cloned 3' to the b1-3 message. These 0.6 kbp obviously originated from a messenger RNA, as they were cloned with a polyA tail. Partial sequence data for this DNA were used for FASTA and BLAST searches, but revealed no significant homology to any sequences in the databases. The partial DNA sequence of this unidentified message cloned in the pb1-3cDNA clone is given below.

Partial DNA sequence of unidentified cDNA cloned in the pb1-3cDNA plasmid (reverse complement of 3' end):

```
TTTTTTTTTT TTTTTTTTTT TTTTAAAC TAATGGGGAC AATTAATTTC
TCGCTGAGGT ACGCTCAGGC CCGCCAGAAC CTGTCGATGG GATACCCGAC
TCTGCAGGTG CCTTGACGGG NCGCTCACCG GCTTNNTTCC ATTCCGAGGT
AGCCCTCNAG TCTCGACGGT AGACGTTACA GTGTTCTTCA ATCCCTGAAG
CATCTCTCCA GGCATGTTGA GGCCGCAGAG GCAGTACTTG GAATGAGCCA
AGGATCCAGT GTT
```

The b2-3 cDNA hybrid clone pb2-3cDNAh contained a non mating type message cloned in front of a partial b2-3cDNA. Partial DNA sequence data were obtained for this message as shown below.

Partial DNA sequence of cDNA for ribosomal protein L27 from the pb2-3cDNAh clone:

```
AAAAAATCCA TCTGGCCTGT CATCTCTCCT GCCGTCGAAC TCATGTCCTT
TCTGACTCGG TGTTGCGCCC AAGCGCGAGG AACCCCTTCT CGCCCTTCGG
CCTCGGGTCT ATTCGTACAG CAACAAAGAG AGCTGGAGGA ACTGTTAACA
ATCATGGTGG TCACCAGGTC AACGGCTGGG TGTAAGAAG TTCTCAGACG
AATATGTGAT CCCTGGCAAC ATTATCGTTA GACAACGCGG GACCTTG TTC
CACCCAGGCC AACACGTCGG AATGGGACGC GACCACACGA TTTTCGCAAC
TGCACCTGGT TATGTGCGCT TCTGGAAGGA GAAGTACATG CGCGGAGAAC
GGAGATACGT TGGCGTGGTG TTGGAGAAGG GAGACAAGCT TCCCTTCGAG
GAATGAAGCC GAGCGAGGTA CCAGCAGATA TTGCCCATTC GTCAACCTCA
ACCACTTCCG CAAAAGGCCA GCAGTTTCTC ATACCTCTC GTACCTCGCG
GTCATGATAG TATCTGATCA TCTATAACGA GATCTGCAGC GAAAAAAAAG
TTATCGTAGA ACGGCCATCG GCTTATCGCT TCTAACCTTT TCTCGACA
```

BLAST searches revealed significant homology to messages for the ribosomal protein L27 (protein 27 of the large subunit of the ribosome) as shown below. These data suggest, that accidentally during this study a partial cDNA sequence for the *C. cinereus* ribosomal protein L27 was obtained. As the similarity is found to prokaryotic and mitochondrial/chloroplast ribosomal protein L27, it is most likely, that the gene cloned from *Coprinus* is the one for the mitochondrial ribosomal protein.

Results of a BLAST search at the NCBI computer for the 5' end of the pb2-3cdnah clone:

		Reading Frame	High Score	Smallest Poisson Probability P(N)
RL27_TOBAC	50S RIBOSOMAL PROTEIN L27, CHLOR...	+3	221	8.7e-25
RL27_ECOLI	50S RIBOSOMAL PROTEIN L27. >pir ...	+3	201	1.4e-22
RM07_YEAST	MITOCHONDRIAL 60S RIBOSOMAL PROT...	+3	203	3.6e-21

>sp|P30155|RL27_TOBAC 50S RIBOSOMAL PROTEIN L27,
CHLOROPLAST PRECURSOR (CL27).

Score = 221 (104.1 bits), Expect = 8.7e-25, P = 8.7e-25
Identities = 42/74 (56%), Positives = 54/74 (72%), Frame = +3

Query:108 SIRTATKRAGGTVNNHGGSPGQRLGVKKFSDEYVIPGNIIVRQRGTLFHPGQHVGMRDHT 287
+I +A K+ G+ N SPGQRLGVK F D+ PG+IIVRQRGTFHFG++VG+G+DH
Sbjct: 48 TIESAHKKGAGSTKNGRDSFGQRLGVKIFGDQVAKPGSIIVRQRGTKFHPGKENVGLGKDH 107

Query:288 TIFATAPGYVRFWK 329
TIF+ G V+F K
Sbjct:108 TIFSLIDGLVKFEK 121

>sp|P02427|RL27_ECOLI 50S RIBOSOMAL PROTEIN L27.
Score = 201 (94.6 bits), Expect = 1.4e-22, P = 1.4e-22
Identities = 40/68 (58%), Positives = 49/68 (72%), Frame = +3

Query:120 ATKRAGGTVNNHGGSPGQRLGVKKFSDEYVIPGNIIVRQRGTLFHPGQHVGMRDHTIFA 299
A K+AGG+ N S +RLGVK+F E V+ G+IIVRQRGTFH G +VG GRDHT+FA
Sbjct: 2 AHKKAGGSTRNGRDSFAKRLGVKRFGGESVLAGSIIVRQRGTKFHAGANVGCGRDHTLFA 61

Query:300 TAPGYVRF 323
A G V+F
Sbjct: 62 KADGKVKE 69

>sp|P12687|RM07_YEAST MITOCHONDRIAL 60S RIBOSOMAL PROTEIN
MRP7 PRECURSOR (YMR6)

Score = 203 (95.6 bits), Expect = 3.6e-21, P = 3.6e-21
Identities = 38/76 (50%), Positives = 55/76 (72%), Frame = +3

Query:111 IRTATKRAGGTVNNHGGSPGQRLGVKKFSDEYVIPGNIIVRQRGTLFHPGQHVGMRDHT 290
+R ATKRA G+ + S G+RLG KK+ + V G II+RQRGTF+PG++VG+G+DH+
Sbjct: 25 VRNATKRAAGSRTSMKDSAGRRLGPKKYEGQDVSTGEIIMRQRGTKFYPGENVGIGKDHS 84

Query:291 IFATAPGYVRFWKEKY 338
IFA PG VR++ + +
Sbjct: 85 IFALEPGVVRYYLDPF 100

Score = 45 (21.2 bits), Expect = 0.83, Poisson P(2) = 0.57
Identities = 10/30 (33%), Positives = 16/30 (53%), Frame = +3

Query:348 ERRYVGVVLEKGDKLPRNEAERGTSRYCPF 437
 +R+++GV L + KLP E R+ F
Sbjct:103 KRKFIGVALRRDLKLPSHFEP TVRRFGRF 132

The important value to assess the quality of the match is the smallest Poisson probability, which indicates the probability that a particular match could be obtained just by chance. This probability is extremely low at 8.7×10^{-25} .

