

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/101217>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

THE BRITISH LIBRARY DOCUMENT SUPPLY CENTRE

TITLE

THE LECTIN GENE FAMILY OF RICINUS COMMUNIS

AUTHOR

James W. Tregear,

INSTITUTION
and DATE

University of Warwick
1989

Attention is drawn to the fact that the copyright of this thesis rests with its author.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no information derived from it may be published without the author's prior written consent.

THE BRITISH LIBRARY
DOCUMENT SUPPLY CENTRE
Boston Spa, Wetherby
West Yorkshire
United Kingdom



REDUCTION X

20

CAMERA 2



THE LECTIN GENE FAMILY OF RICINUS COMMUNIS

James W. Tregear, BSc (Hons) Edinburgh

A thesis submitted for the degree of
Doctor of Philosophy
at the University of Warwick

Plant Biochemistry Group
Department of Biological Sciences
University of Warwick
Coventry CV4 7AL
United Kingdom

August, 1989

Dedicated to

Mum, Dad and Uncle Leathan

CONTENTS

	PAGE
List of Contents	i
List of Tables	ix
List of Figures	x
Acknowledgements	xii
Declaration	xiii
Summary	xiv
Abbreviations	xv
SECTION I. LITERATURE REVIEW	1
I.1. INTRODUCTION	1
I.2. GENERAL FEATURES OF PLANT GENE STRUCTURE	4
I.2.A. Promoter sequences necessary for the transcription of nuclear genes	4
I.2.B. mRNA 5' leader sequences	10
I.2.C. mRNA 3' untranslated regions	11
I.2.D. Intervening sequences	13
I.2.E. Initiation and termination of translation	14
I.2.F. Signal and linker peptide sequences	15
I.2.G. Pseudogenes	17
I.3. <u>Ricinus communis</u> lectins and their biogenesis	20
I.3.A. General features of castor bean seed development	20

	PAGE
I.3.B. General properties of castor bean lectins	22
I.3.C Synthesis and processing of castor bean lectins	28
I.3.D. Primary structure and folding of castor bean lectins	30
I.3.E Structure and organisation of castor bean lectin genes	34
I.4. DEVELOPMENTAL AND TISSUE-SPECIFIC EXPRESSION OF PLANT GENES	41
I.4.A. Analysis of plant gene expression	41
I.4.B. Heterologous gene expression as a tool for analysing <u>cis</u> -acting regulatory sequences	42
I.4.C <u>Trans</u> -acting DNA binding proteins	46
I.4.D Identification of regulatory sequences and factors	48
I.5. MOLECULAR CLONING AND ITS APPLICATION TO PLANT SYSTEMS	57
I.5.A. Principles and strategies	57
I.5.B. cDNA cloning	63
I.5.C. Nucleotide sequence analysis	64
I.5.D Genomic cloning	66
I.6. AIMS AND APPROACHES	74
SECTION II. MATERIALS AND METHODS	75
II.1. MATERIALS	75
II.1.A. Plant material	75

	PAGE
II.1.B. Chemicals, biochemicals, radiochemicals and enzymes	75
II.2 GROWTH AND STORAGE OF BIOLOGICAL MATERIALS	77
II.2.A. Growth of plants	77
II.2.B. Growth and storage of bacterial stocks	78
II.2.C. Growth and storage of bacteriophage lambda strains	78
II.2.D. Growth and storage of M13 bacteriophages	81
II.3. ISOLATION OF CASTOR BEAN NUCLEAR DNA	82
II.3.A. Isolation of nuclei	82
II.3.B. Isolation of DNA from castor bean nuclei	84
II.3.C. CsCl gradient purification of DNA	85
II.4. SOUTHERN BLOTTING AND HYBRIDISATION OF GENOMIC DNA	87
II.4.A. Restriction endonuclease digestion	87
II.4.B. Agarose gel electrophoresis	87
II.4.C. Transfer of DNA onto hybridisation membranes	88
II.4.D. Hybridisation of radiolabelled probe to immobilised DNA	90
II.5. CONSTRUCTION OF GENOMIC LIBRARIES IN BACTERIOPHAGE LAMBDA	93
II.5.A. Determination of optimal conditions for the production of 10-22 Kb genomic fragments	93
II.5.B. Large scale preparation of genomic fragments for cloning	94

II.5.C.	Fractionation of genomic fragments	PAGE 95
II.5.D.	Preparation of bacteriophage lambda arms	97
II.5.E.	Ligation of genomic fragments to lambda arms	99
II.5.F.	Packaging of ligated DNA and titration of bacteriophages obtained	100
II.6.	SCREENING OF LAMBDA GENOMIC LIBRARIES FOR LECTIN SEQUENCES	101
II.6.A.	Growth of lambda plates for screening	101
II.6.B.	Plaque lifts	101
II.6.C.	Plaque hybridisation	102
II.7.	PREPARATION AND ANALYSIS OF DNA FROM POSITIVE CLONES	105
II.7.A.	Growth of bacteriophages and extraction of DNA	105
II.7.B.	Restriction and Southern blotting analysis of DNA	111
II.8.	SUBCLONING OF LECTIN GENES AND THEIR FLANKING SEQUENCES INTO PLASMID VECTORS	112
II.8.A.	Restriction and dephosphorylation of plasmid vector	112
II.8.B.	Gel isolation of DNA fragments	113
II.8.C.	Ligation of DNAs	114
II.8.D.	Transformation of ligated plasmids into <u>E. coli</u>	115
II.9.	PREPARATION OF PLASMID DNA	117
II.9.A.	Rapid, small scale preparations	117
II.9.B.	Large scale preparations	118

II.9.C.	Purification of plasmid DNA on CsCl gradients	PAGE 120
II.10.	RESTRICTION MAPPING OF PLASMID SUBCLONES	121
II.11.	DIDEOXYNUCLEOTIDE SEQUENCING OF LECTIN CLONES	122
II.11.A.	Subcloning of DNA fragments into M13 vectors	122
II.11.B.	Preparation of single-stranded DNA templates	123a
II.11.C.	Sequencing reactions	123b
II.11.D.	Polyacrylamide gel electrophoresis of sequencing products	126
II.12.	ISOLATION OF RNA FROM CASTOR BEAN TISSUES	128
II.12.A.	Isolation of total RNA from castor bean endosperm	128
II.12.B.	Isolation of total RNA from castor bean roots and leaves	129
II.13.	NORTHERN BLOTTING AND HYBRIDISATION	131
II.13.A.	RNA gels	131
II.13.B.	Northern transfer and hybridisation	131
II.14.	ANALYSIS OF IN VIVO TRANSCRIPTION OF LECTIN GENES USING RNASE PROTECTION	133
II.14.A.	Construction and linearisation of plasmids for transcription	133
II.14.B.	Preparation of radiolabelled RNA probes	133
II.14.C.	RNase protection	135
II.15.	RADIOACTIVE LABELLING OF DNA <u>IN VITRO</u>	138

II.15.A.	Nick translation of DNA fragments	PAGE 138
II.15.B.	Oligolabelling of DNA fragments	139
II.15.C.	Radilabelling of DNA fragments using an end filling reaction	140
SECTION III. RESULTS		142
III.1.	GENOMIC ORGANISATION OF CASTOR BEAN LECTIN GENES	142
III.1.A.	Estimation of <u>R. communis</u> genome size	142
III.1.B.	Estimated number of genes in the <u>R. communis</u> genome	142
III.2.	NUMBER AND DIVERSITY OF CLONED LECTIN GENES	147
III.2.A.	Construction and screening of genomic libraries	147
III.2.B.	Southern analysis of lectin-positive clones	155
III.3.	DETAILED RESTRICTION ANALYSIS OF SELECTED CLONES	159
III.4.	SELECTIVE NUCLEOTIDE SEQUENCE ANALYSIS OF CLONES	167
III.5.	LEVELS AND PATTERNS OF LECTIN GENE TRANSCRIPTS	181
III.5.A.	Isolation of total RNA	181
III.5.B.	Northern analysis of lectin gene transcripts	183
III.5.C.	RNAse protection analysis of transcripts using probes from the PCBG3H1 (lambda clone 10) ricin gene	183

	PAGE
III.5.D. RNAse protection analysis of transcripts using internal probes from the putative lectin genes in pCBG2H1 (lambda clone 3) and pCBG4H2 (lambda clone 7)	187
SECTION IV. DISCUSSION	193
IV.1. DIVERSITY OF LECTIN CLONES	193
IV.1.A. Restriction site diversity	193
IV.1.B. Nucleotide sequence diversity	198
IV.1.C. Size and quality of genomic libraries	208
IV.2. IN VIVO TRANSCRIPTION OF CASTOR BEAN LECTIN GENES	216
IV.2.A. Northern analysis of transcripts	216
IV.2.B. Verification of transcriptional inactivity of the pCBG2H1 (lambda clone 3) and pCBG4H2 (lambda clone 7) pseudogenes	217
IV.2.C. Analysis of in vivo transcripts of the pCBG3H1 (lambda clone 10) ricin gene	220
IV.3. POSSIBLE TRANSCRIPTIONAL REGULATORY SIGNALS IN THE pCBG3H1 (LAMBDA CLONE 10) RICIN GENE PROMOTER	234
IV.3.A. Identification of TATA box(es)	234
IV.3.B. Identification of a putative CAAT/ACGA box analogue	235
IV.3.C. Identification of sequences which might confer seed-specific expression	237
IV.4. SIZE AND EVOLUTIONARY RELATIONSHIPS OF THE CASTOR BEAN LECTIN GENE FAMILY	241

	PAGE
IV.4.A. Estimated size of the lectin gene family	241
IV.4.B. Linkage relationships	244
IV.4.C. Possible evolutionary sequence in the formation of the lectin gene family	246
IV.5. FUTURE GOALS AND FURTHER APPLICATIONS OF CLONED LECTIN GENES	251
IV.6. CONCLUDING REMARKS	255
APPENDIX I:	
Restriction maps of vectors used	256
REFERENCES	259

LIST OF TABLES

<u>Table</u>	<u>Title</u>	<u>PAGE</u>
1	Chemical treatments used for Maxam and Gilbert sequencing	68
2	Bacteria, bacteriophages and plasmids used	79
3	Composition of working solutions for sequencing	125
4	Estimation of <u>E. communis</u> C-value from nuclei counts and DNA yields	143
5	Yields of bacteriophages from lambda <u>in vitro</u> packaging reactions	153
6	Classification of lambda lectin-positive clones	158
7	M13 subclones constructed for sequencing	169
8	Oligonucleotides used for M13 dideoxy-sequencing	170
9	Clones constructed for RNase protection analysis	186

x LIST OF FIGURES

Figure	Title	PAGE
1	Generalised structure of a plant nuclear gene transcribed by RNA polymerase II.	5
2	Principle of the chain termination sequencing method	65
3	Principle of the Maxam and Gilbert sequencing method	67
4	Genomic cloning strategies; construction of a castor bean genomic library	70
5	Identification of genomic DNA restriction fragments containing lectin-like sequences	145
6	Identification of genomic DNA restriction fragments sharing high homology with a ricin cDNA probe	146
7	Assessment of optimal conditions for partial digestion of castor bean DNA	150
8	Size analysis of digested genomic DNA fractions removed from a sucrose gradient	152
9	Southern blotting analysis of lectin-positive lambda clones (1-9)	157
10	Southern blotting analysis of lectin-positive lambda clones (10-17)	157
11	Restriction maps of pCBG1R2 and pCBG1R5	161
12	Restriction map of pCBG2H1	161
13	Restriction map of pCBG3H1	162
14	Restriction maps of pCBG4H2 and pCBG4H4	165
15	Comparison of restriction sites in the lectin gene subclones and outline of sequencing and RNase protection strategies	166

	PAGE
16 Nucleotide sequence of pCBG3H1 (lambda clone 10) ricin gene	172
17 Partial nucleotide sequence of pCBG1R2/pCBG1R5 (lambda clone 1) lectin pseudogene	175
18 Nucleotide sequence of an internal region of the pCBG2H1 (lambda clone 3) lectin pseudogene	179
19 Nucleotide sequence of an internal region of the pCBG4H2 (lambda clone 7) lectin pseudogene	180
20 Stages of <u>R. communis</u> seed development	182
21 Northern analysis of lectin transcript levels in different tissues and stages of seed development	184
22 RNase protection analysis of transcripts using a probe spanning the 5' end of the pCBG3H1 (lambda clone 10) ricin gene	188
23 RNase protection analysis of transcripts using a probe spanning the 3' end of the pCBG3H1 (lambda clone 10) ricin gene	189
24 RNase protection analysis of transcripts using an internal probe from the putative gene in pCBG2H1 (lambda clone 3)	191
25 Predicted cap sites of the pCBG3H1 (lambda clone 10) ricin gene	227
26 Predicted poly(A) sites of the pCBG3H1 (lambda clone 10) ricin gene	232
27 Comparison of the pCBG3H1 (lambda clone 10) ricin gene and castor bean 2S albumin gene promoter sequences	236
28 Identification of a sequence motif conserved in the promoters of the pCBG3H1 (lambda clone 10) ricin gene and gcepla 2S albumin gene	239
29 Possible evolutionary sequence in the formation of the lectin gene family	247

ACKNOWLEDGEMENTS

I would like firstly to thank my supervisor, Dr Lynne Roberts, for her valued advice and assistance, both technical and general, during the course of this project. Thanks are also owed to Prof Mike Lord for making a number of things possible and I would also like to thank my colleagues in the ricin group for their cheerful repartee whilst I was working with them. I have been fortunate to receive technical help from a large number of people in the Plant Biochemistry Group, too numerous to list entirely, but I would nevertheless like to thank for help with specific techniques (in alphabetical order) Chris Bowler, Mark Cock, Phil Gilmartin, Dave Griffiths, Steve Irwin, Ann Smith and Robert Spooner. Also in connection with the experiments described in this thesis I would like to thank Mr Matt Busby for growing the castor bean plants and Mr Dick Barker for the Maxam and Gilbert sequencing data. I acknowledge the financial support of the SERC. Finally, I would like to gratefully thank my family for their support over the last three or more years. I am particularly grateful to Kata for typing this thesis and to Jane and Steve, who put me up whilst I was writing it. I would also like to thank Betty and Dave for bearing with me during the later, most chaotic stages and Anthony Williams & Co for providing photocopying facilities.

DECLARATION

All the results presented in this thesis were obtained by the author, apart from the Maxam and Gilbert sequencing data, which was obtained by R Barker, as indicated in the text.

All sources of information have been specifically acknowledged by means of reference. None of the work contained in this thesis has been used for any previous application for a degree.

James W. Tregear

James W. Tregear

SUMMARY

The aim of this project has been to isolate, characterise and investigate the expression patterns of the various members of the castor bean lectin gene family. Genomic Southern blotting experiments showed that the gene family contains approximately eight members, of which two appear to be ricin-like.

Castor bean genomic DNA was cloned into a bacteriophage lambda vector and the resulting recombinant clones screened with a ricin cDNA probe. Seventeen positive clones were isolated, amongst which at least five different groups could be recognised, on the basis of Southern blotting data. Four different lambda clones were selected for further analysis. One of the clones was found to contain a functional ricin gene with an identical restriction pattern to that of the ricin cDNA, but none of the clones appears to contain an authentic Ricinus communis agglutinin (RCA I) gene. DNA sequencing and RNase protection data showed that three of the clones analysed contain lectin pseudogenes.

The expression pattern of the functional ricin gene in pCBG3H1 (lambda clone 10) was investigated at the transcriptional level using RNase protection. The results obtained show that the mRNA transcribed from this gene accumulates during the late (post-testa) stages of seed development. The pCBG3H1 ricin gene appears to use multiple cap and poly(A) sites.

The developmental profile of lectin gene transcript levels observed in this study is similar to the patterns previously observed at the protein and translatable mRNA levels. This indicates a close correlation between the accumulation of the lectin proteins and the amounts of steady state transcripts. DNA sequencing enabled the identification of putative transcriptional regulatory elements in the promoter of the pCBG3H1 ricin gene, including an element resembling the RY repeat previously implicated in seed-specific gene expression.

Ultimately, it is hoped that studies of this type will make it possible for the network of regulatory processes governing the expression pattern of the castor bean lectin genes to be unravelled.

ABBREVIATIONS

A, ATP, dATP, ddATP	adenine, adenosine triphosphate, deoxyadenosine triphosphate, dideoxyadenosine triphosphate
bp	base pair
cDNA	complementary DNA
Ci	curie
cm	centimetre
C, CTP, dCTP, ddCTP	cytosine, cytidine triphosphate, deoxycytidine triphosphate, dideoxycytidine triphosphate
D	dalton
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
EDTA	ethylenediaminetetra-acetate
g	gramme
G, GTP, dGTP, ddGTP	guanine, guanosine triphosphate, deoxyguanosine triphosphate, dideoxyguanosine triphosphate
h	hour
IPTG	Isopropylthiogalactoside
Kb	kilobase
KD	kilodalton
l	litre

m	metre
min	minute
ml	millilitre
N, NTP, dNTP, ddNTP	nucleotide, nucleoside triphosphate, deoxynucleoside triphosphate, dideoxynucleoside triphosphate
pfu	plaque forming unit
PEG 6000	polyethylene glycol of molecular weight 6000 D
RCA I	<u>Ricinus communis</u> agglutinin I
RNA	ribonucleic acid
RNAse	ribonuclease
rpm	revolutions per minute
rubisco	Ribulose 1,5-bisphosphate carboxylase-oxygenase
SDS	sodium dodecyl sulphate
SV40	simian virus 40
TEMED	N,N,N',N'-tetramethyl-ethylene- diamine
tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
T, TTP, dTTP, ddTTP	thymine, thymidine triphosphate, deoxythymidine triphosphate, dideoxythymidine triphosphate
v/v	volume per volume
w/v	weight per volume

X-gal

5-bromo,4-chloro-3-indolyl- β -D-
galactoside

ug

microgramme

ul

microlitre

SECTION I

LITERATURE REVIEW

SECTION I.1. INTRODUCTION

In the late 19th century, the Austrian monk Gregor Mendel postulated that the inheritance of specific traits from one generation of an organism to the next could be considered in terms of finite genetic units. These units, which he referred to as "factors" have since become commonly known as genes. Initially the discovery of genes made it possible for the process of heredity to be studied in a scientifically meaningful way. Subsequently, classical genetics has been used to identify and characterise a host of important genetic phenomena. Nevertheless the study of genetics was revolutionised in 1944 when Avery et al presented evidence that DNA was the molecule which carried the genetic information contained in cells. This discovery made it possible for genetics and biochemistry to be used in tandem to provide a new insight into the structure and function of genes. Thus whilst many of the early discoveries concerning DNA were achieved using chemical and physical techniques, there has since been an explosion of research in a relatively new area known as molecular biology. This discipline is concerned in particular with the isolation of genes by molecular cloning using recombinant DNA technology, the determination of DNA sequences, and the manipulation of DNA molecules in vitro. By using such an approach, the

study of gene structure and function has been revolutionised. It is hoped that work of the type described in this project will ultimately contribute towards the elucidation of the mechanisms by which higher plant genes are "switched on and off" during the course of development. A large number of higher plant species are known to be totipotent, ie capable of regenerating into an intact mature plant from a single living cell. Regeneration can normally be achieved with both differentiated and undifferentiated cell types. It therefore seems reasonable to assume that, in general, plant cells carry their entire genetic complement throughout development. Thus, a complex system for gene regulation must exist in order that differentiation and development can be successfully coordinated.

The general approach of molecular biology has been to work "from the ground up", ie to isolate and characterise genes, identify DNA elements potentially involved in their regulation, and then to look for factors in vivo which may interact with the DNA sequences of interest. Ultimately it would be hoped that a network of regulatory mechanisms might be identified through studies on a large number of genes. Unfortunately the cloning of genes involved in the regulation of development in general would probably be difficult as they would be expected to have mRNA and protein products of very low abundance. Thus although

some newer techniques may circumvent this problem, to date most of the plant genes isolated have been those expressed at relatively high levels. In particular many of the genes coding for proteins involved in photosynthesis have been cloned and studied, as have many seed storage protein genes. The lectin gene family of Ricinus communis L. which was the system investigated in this study, falls into the latter group. The isolation and characterisation of the genes involved is a first step towards understanding how their regulation might be achieved.

SECTION 1.2. GENERAL FEATURES OF PLANT GENE STRUCTURE

Molecular cloning techniques have been successfully applied to a wide range of plant systems and a large quantity of data on plant gene structure has now been amassed. Fig 1 illustrates the essential features of a generalised plant gene.

1.2.A. Promoter sequences necessary for the transcription of nuclear genes

The transcriptional activity of plant genes is regulated by cis-acting DNA elements located in the gene flanking regions. The cis-acting elements of a gene promoter may be broadly classified in functional terms into two types, namely those elements which exert a quantitative effect on transcription and those which exert a qualitative effect. The latter elements, which are responsible for determining the developmental and tissue-specific patterns of the genes upon which they act, will be discussed in section 1.4. As will also be discussed, the original distinctions made between these two types of element now appear to be something of an oversimplification. This may reflect in part, however, the considerable difficulties associated with dissecting the functions of a promoter into its component parts. The discussion in this section will be

Figure 1

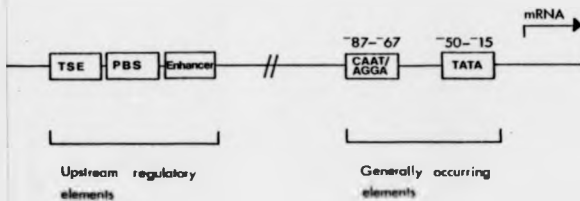


Figure 1 Generalised structure of a plant nuclear
gene transcribed by RNA polymerase II

All plant nuclear genes which encode proteins are transcribed by RNA polymerase II. The TATA box shown is thought to act as a binding site for this enzyme. The position of both the TATA and CAAT (or AGGA) boxes seems to be similar in all genes of this type. In contrast, the position of the upstream regulatory elements shown appears to vary considerably amongst different genes. The configuration shown in the diagram is therefore purely arbitrary. In many cases, one or more of the upstream elements may be present in multiple copies.

The box marked TSE represents those elements which have been shown by functional analysis to be involved in the tissue and/or developmental specificity of plant gene promoters. The box marked PBS represents those elements which have been identified as trans-acting protein binding sites. Finally, a third box represents those elements which have been shown to exhibit enhancer-like activity. The representation of these three types of element as individual boxes is intended mainly as a functional classification, since in many cases their respective activities have not been resolved by genetic dissection.

restricted to those DNA elements which exert a quantitative effect on transcriptional levels.

I.2.A.i) The TATA or Goldberg Hogness box

This regulatory element, so called because of its core consensus sequence (Proudfoot, 1979), is normally found approximately 25 bp upstream from the cap site (the point corresponding to the 5' end of the mRNA) of both plant and animal genes, although its exact location may vary. The TATA sequence is thought to act as a recognition site for RNA polymerase II, the enzyme responsible for the transcription of nuclear genes which code for proteins. The TATA box is therefore involved in determining the site of transcription initiation. Although the consensus sequences for plant and animal TATA boxes vary slightly, their core sequences are nevertheless the same.

I.2.A.ii) The AGGA or CAAT box

Most animal gene promoters have been found to contain a CAAT box (again so called because of its consensus sequence) further upstream from the TATA box. CAAT boxes usually occur approximately 80-100 bp upstream from the cap site, although again this is only a general rule. It has been shown, at least in some cases, that the transcriptional activity of a gene may be significantly reduced when the CAAT box sequence is deleted (Grosveld et al., 1982). Nevertheless the exact mechanistic role of

this element remains undetermined. DNA elements similar to the CAAT box have been found in the promoter regions of a large number of plant genes. In many cases, however, a related element is seen to occur in place of the typical CAAT box. This element has become known as the AGGA box (Messing et al, 1983). By comparing the DNA sequences of various plant promoters, Messing et al (1983) assigned to the AGGA box the consensus sequence C(C/T)A(2-5)(G/T)NGA(2-4)(C/T)(C/T). It can be seen that the CAAT sequence seen in some plant promoters is compatible with this consensus sequence.

I.2.A.iii) Enhancer-like elements

Enhancer elements were first discovered in the promoter regions of genes cloned from animal viruses. The first element of this type to be characterised was that found in the early transcriptional control region of the SV40 virus. This enhancer consists of a 72 bp tandem repeat located more than 150 bp upstream from the cap site of the early viral genes (Benoist and Chambon 1981). In common with the two types of element previously described, the SV40 enhancer exerts a positive effect on transcriptional levels. This stimulatory effect was, however, found to be relatively independent of both its position and orientation. Banerji et al (1981) showed that the SV40 enhancer remained active when positioned 1400 bp upstream or 3300 bp downstream from the transcription initiation

site of a rabbit beta-globin gene introduced into Hela cells. Moreover, the enhancer was shown to be active in both orientations. Subsequent work has shown that enhancers can act over considerable distances, sometimes as large as 10 kb (Khoury and Gruss, 1983). The most important criterion appears to be the sequences located between the enhancer element and promoter of the gene being assayed, as some DNA sequences are able to interfere with the functioning of enhancers. Since the discovery of the SV40 enhancer, elements with similar properties have been discovered in the genomes of various other viruses such as adenoviruses and papillomaviruses, as well as in animal genomes themselves. Enhancer-like elements have also been detected in the promoter regions of a number of plant genes. One of the first plant enhancer-like sequences to be described was that of the cauliflower mosaic virus 35S promoter (Odell *et al.*, 1985). By examining the effects of deletions on the *in vivo* activity of the promoter, it was shown that the region between -46 and -105 (with respect to the mRNA cap site) contains sequences which greatly increase the level of transcription in a similar manner to the enhancer elements of animal systems. The region in question includes a CAAT box and an inverted repeat region containing an SV40 enhancer core consensus sequence. Odell *et al.* (1988) showed that the activity of the 35S promoter was reduced when the CAAT box, but not the core enhancer sequence, was

deleted. Another deletion retaining the CAAT box but only a portion of the inverted repeat showed some activity, but this was not sufficient to account for the loss in activity observed in the absence of the CAAT box. It therefore appears that the intact inverted repeat itself plays a role in transcription stimulation. Further evidence for this conclusion was provided by the observation that a 182 bp fragment containing both copies of the SV40 virus 72 bp repeat showed no enhancer activity when placed behind a nopaline synthase (nos) promoter at -150.

Other studies have shown that the activities of enhancer-like elements in plant gene promoters may be of a developmental or tissue-specific nature (eg Chen *et al.*, 1988). Furthermore, some plant enhancer elements have also been shown to act as tissue-specific silencers (Simpson *et al.*, 1986). It can therefore be seen that considerably more work must still be carried out to elucidate the structure and mode(s) of action of these elements. Sequences resembling the animal enhancer core consensus sequences have in fact been found in the promoter regions of a large number of plant genes with diverse functions. Examples include several pea legumin genes (Lycett *et al.*, 1985), various maize zein genes (Roussell *et al.*, 1988), a sweet potato sporamin gene (Hattori and Nakamura, 1988) and a tobacco chlorophyll a/b

binding protein gene (Castresana et al, 1988). In many cases (such as the latter), the enhancer-like sequences are present as multiple repeats.

I.2.B. mRNA 5' leader sequences

As the name implies, the 5' untranslated region or 5' leader sequence of a gene is the transcribed region which precedes the translation initiation codon. Following transcription, at least some plant nuclear mRNAs are modified at their 5' terminus to produce a 5' cap structure similar to that observed in animal systems. The capping reaction involves the addition of a 7-methylguanine residue via a 5'-5' triphosphate linkage (Grierson, 1982). In some cases, one or more nucleotides may be removed from the 5' end of the mRNA prior to capping. For this reason, the cap site may not always correspond to the transcription start site. Although the mRNA 5' leader sequence is not itself translated, it is clear that it plays an important role in determining the translational efficiency of the molecule. The most striking example is the 5' leader sequence of the tobacco mosaic virus mRNA, which has been shown to enhance the translation of contiguously linked foreign gene transcripts both in vitro and in vivo (Gallie et al, 1987). In some cases, it has been observed that the translational efficiency of certain mRNA species appears

to change during development (Hughes and Galau, 1987, Fabijanski and Altosaar, 1985). It therefore seems likely that the phenomena observed are governed in some way by interactions between the cellular translational apparatus and the 5' leader sequences of the mRNAs in question.

I.2.C. 3' untranslated regions

Most, if not all plant nuclear mRNAs are post-transcriptionally polyadenylated at their 3' ends. The information necessary to direct the addition of the poly (A) tail, which may be up to 200 nucleotides in length, is thought to reside in the 3' untranslated region of the precursor mRNA. Unfortunately the polyadenylation process of plants is less well characterised than that of animal systems. However, it seems that the process is generally similar in both. The polyadenylation process may be subdivided into two stages. In the first stage the precursor mRNA is endonucleolytically cleaved at a specific point in the 3' untranslated region. The point of cleavage is usually referred to as the poly(A) site. The second stage of the process involves the addition of adenine residues to the cleaved end. In animal systems the position of the poly(A) site is thought to be determined by polyadenylation signals both upstream and downstream from the cleavage point (Proudfoot, 1984). A polyadenylation signal consisting of the hexanucleotide

AATAAA has been detected 10 to 30 bp upstream from the poly(A) site in all animal nuclear mRNAs. It has been shown that abnormal or inefficient 3' processing results when this sequence is mutated or deleted (Wickens and Stephenson, 1984, Fitzgerald and Shenk, 1981).

Furthermore, it has been suggested that the hexanucleotide sequence interacts in some way with small nuclear ribonucleoproteins (snRNP) to bring about cleavage site selection (Hashimoto and Staltz, 1986). Whilst it is generally assumed that much of the information obtained on polyadenylation in animal systems may be extrapolated to plant systems, it appears nevertheless that some differences do occur. Joshi (1987) examined the nucleotide sequences obtained from 46 different plant genes with a view to establishing a consensus sequence for plant polyadenylation signals. A full discussion is not appropriate here, but three points should be noted:

i) A consensus AATAAA sequence (or a close approximation thereof) is present in most plant 3' untranslated regions, typically 10-50 bp upstream from the poly(A) site. Approximately half the genes examined contained a single base substitution (usually in the sixth position), but only 3 genes showed 2 mismatches.

ii) The nucleotide sequences under investigation were scanned for a CATTG consensus sequence commonly observed

in animal pre-mRNAs. It has been suggested that this sequence may facilitate the binding of a U4 snRNP potentially involved in 3' cleavage (Bergert, 1984). Interestingly, a similar motif of the consensus sequence YAYTG was commonly observed, but in varied positions both upstream and downstream from the AATAAA signal. This suggests that the element may not play the functional role previously suggested.

iii) The nucleotide sequences were scanned for a third consensus sequence YGTGTTY which is located approximately 30bp downstream from the AATAAA signal in a large proportion of animal genes. It has been suggested that this element plays a role in determining cleavage specificity (McLaughlan et al., 1985). Homologies with this element were found throughout the 50bp region immediately 3' to the AATAAA signal as well as around the poly (A) site. However, in view of the considerable overlap observed amongst sequences of this type, no functional significance could be inferred.

I.2.D. Intervening sequences

In addition to capping and polyadenylation, a large number of plant nuclear mRNAs undergo a third post-transcriptional modification, commonly known as splicing. Many plant genes, in common with animal genes, contain

non-coding sequences which interrupt the reading frame of the encoded polypeptide. Although these non-coding sequences are initially transcribed, they must be removed from the mRNA prior to translation in order to create an uninterrupted reading frame for the cytosolic ribosomes. A number of small nuclear RNAs (snRNAs) are thought to be involved in the splicing process (Hanley and Schuler, 1988). Intervening sequences, also known as introns, are invariably bordered by the dinucleotide sequence GU at the 5' splice site and AG at the 3' splice site. This sequence information is, however, insufficient in itself to direct the splicing process. Furthermore, it appears that different phylogenetic groups use different consensus sequences for splicing. This is apparent when intron-containing genes are expressed in heterologous systems. Phylogenetic differences in splice site selectivity have been observed to occur both between plants and animals (Van Santen and Spritz, 1987) and between monocotyledonous and dicotyledonous plants (eg Keith and Chua, 1986). Hanley and Schuler (1988) found that three different classes of plant intron could be differentiated by the relative purine or pyrimidine content of sequences upstream from the 3' splice site. These three different types of intron were found to occur at significantly different frequencies in monocots and dicots.

1.2.E. Translation initiation and termination signals

The nuclear genes of plants encode polypeptides by means of the universal genetic code (Khorana, 1968). They therefore use an AUG triplet as the translation initiation codon in the mRNA. According to the scanning model of Kozak (1980), the ribosomal 40S subunit of eukaryotic cells binds, along with its associated factors, to the most 5' AUG triplet on the mRNA, thereby initiating translation. This model appears to be generally applicable to plant systems. Plant nuclear mRNAs use three different codons for translation termination, namely UGA, UAA and UAG. Some plant genes have been found to contain two tandemly linked termination codons (eg Roberts et al, 1985)

I.2.F. Signal and linker peptide sequences

Many plant proteins are synthesised initially as precursors which must be post-translationally modified in order to attain biological activity. In some cases the precursor polypeptide must undergo endoproteolytic cleavage, which may be either coupled to its synthesis (co-translational) or independent of it (post-translational). In general, co-translational processing is typical of polypeptides which are secreted into the endoplasmic reticulum of cells, whereas post-translational cleavage is typical of polypeptides destined for organelles. In many cases, proteolytic

cleavage forms part of the process by which polypeptides are "addressed" to specific locations in the cell. The destination of such polypeptides is normally determined by peptide sequences present at the N-terminus of the precursor molecule. These sequences are commonly known as signal or transit peptides. The length of signal peptides varies considerably, ranging from 20 amino acids in the case of a zein precursor (Messing et al, 1983) to 47 amino acids in the case of a wheat rubisco small subunit precursor (Broglie et al, 1983). In general, it appears that signal peptides in themselves contain sufficient information to target proteins to their correct destinations (Van den Broeck et al, 1985).

Whilst proteolytic processing is an essential and integral part of the intracellular protein sorting mechanism, it is also necessary where the subunits of a protein are synthesised as a single polypeptide precursor. Many plant proteins are synthesised in this way, so their biosynthetic pathways must include an endoproteolytic cleavage step at some point. Some proteins, such as pea legumin (Lycett et al, 1984) and soybean glycinin (Negoro et al, 1985) have been shown to arise from a precursor containing subunit peptides which are contiguously linked. In such cases the protein subunits may be separated by cleavage at the junction between the two respective chains. A number of other plant proteins have been shown

to arise from precursors containing extra peptide sequences which link the subunit molecules. These so-called linker peptides, which usually contain only a small number of amino acid residues, are removed during the biosynthetic pathway and therefore do not appear in the mature protein. Examples of proteins synthesised as precursors containing a linker peptide include the lectins of castor bean (Roberts et al., 1985) and of pea (Higgins et al., 1983).

I.2.G. Pseudogenes

Southern blotting and genomic cloning data has shown that many plant and animal genes are present in multiple "copies" which are usually non-identical. Each group of cross-hybridising genes is normally referred to as a gene family. The individual members of a gene family are often expressed at vastly different levels (eg Voelker et al., 1986). In more extreme cases, it has been shown that some members of a gene family are non-functional. A striking example is the lectin gene family of pea, which has been shown to contain only one functional member (Kaminski et al., 1987). The term "pseudogene" is normally used to describe any non-functional members in a gene family. Pseudogenes may be detected in a number of ways. Firstly, the lack of expression may be due to a cryptic mutation in the 5' flanking region which renders the promoter

inactive. In such cases nuclease protection must be used to demonstrate that no transcripts sharing 100% homology with the genomic clone are synthesised in vivo. In other cases, it may be possible to infer that a gene is non-functional from the nucleotide sequence of the coding region. Frameshift mutations which would lead to premature translation termination have been shown to occur in many pseudogenes. Finally, it has been shown that some pseudogenes lack complete 5' or 3' termini and are therefore equally incapable of coding for a functional protein.

Although pseudogenes are often identified by their lack of transcriptional activity, it is interesting to note that a transcriptionally active plant pseudogene has been isolated. O'Neal et al (1987) described a gene-like DNA element which shared homology with rubisco small subunit genes and was transcribed at a low level, yet could not encode a functional protein since a frameshift mutation was present in the signal peptide sequence.

Pseudogenes may be fundamentally classified into two groups (Vanin, 1984). The first group, known as unprocessed pseudogenes, consists of those non-functional genes which appear to have become inactive by mutation alone. This includes those pseudogenes which have arisen by the insertion of an unrelated piece of DNA such as a

transposable element (Goldberg et al, 1983). The pseudogenes in the second group are composed of gene-like structures which appear to have arisen by the insertion into the genome of reverse transcribed mRNA species. Thus they can be identified by their mRNA-like features, which include the absence of introns and the presence of a poly(A) tract corresponding to the poly(A) tail of the mRNA. Pseudogenes of this type are known as processed pseudogenes and are common in animal systems (eg Vanin et al, 1980). Only one processed pseudogene has been identified in plant systems, however. Drouin and Dover (1987) have described a processed actin pseudogene cloned from potato. It was suggested that the general rarity of processed pseudogenes in higher plants might be due to their lacking an independent and restricted germ line which, if present, might be expected to preserve any pseudogenes arising in this way. It seems, nevertheless, that processed pseudogenes could theoretically arise from any genes which are expressed in the apical meristems of higher plants.

SECTION I.3. RICINUS COMMUNIS LECTINS AND THEIR
BIOGENESIS

I.3.A. General features of castor bean seed development

The castor bean or castor oil plant Ricinus communis L. (family Euphorbiaceae) produces endosperm rich seeds enclosed within a spiny pericarp. Seed maturity is normally attained approximately 12 weeks after pollination, depending on growth conditions. In common with many other species, the storage proteins of castor bean seeds occur in spherical organelles called protein bodies. The single membrane delimiting these structures has been shown by light microscopy to originate from small subvacuolar structures (Gifford et al., 1982). The protein bodies of castor bean seed endosperm consist of an amorphous protein matrix within which two types of inclusion occur. One group of inclusions consists of phytin globoids (Altschul et al., 1966), the other of 11S protein crystalloids (Ory et al., 1968). The 11S crystalloid group form the major protein component of castor bean seeds. They constitute approximately 70 % of the total protein of the mature endosperm (Gifford et al., 1982) and resemble globulins in that they are insoluble in water but soluble in weak salt solutions.

The matrix of the protein bodies consists mainly of two groups of water soluble proteins, the 2S albumins and the 7S lectins. The 2S albumins are the most abundant of these two groups and contribute approximately 20% of the mature endosperm protein (Gifford et al, 1982), whereas the lectins constitute collectively 5.6% of the mature seed protein (Harley and Beevers 1986). The developmental profile of synthesis for both of these groups of proteins appears to be similar (Gifford et al, 1982), with rapid accumulation taking place between 30 and 40 days after pollination. During this period, which coincides with testa formation (Roberts and Lord, 1981A), rapid formation of the protein bodies is also observed.

Although the protein bodies and their constituent proteins are characteristic of the endosperm tissue of castor bean seeds, their presence has also been observed in the cotyledons. Harley and Beevers (1986) detected lectins in the cotyledons of dry castor bean seeds, although the amounts were less than 1% of those in the endosperm. The photosynthetic cotyledons of 10 day old seedlings which had been exposed to light from day 5 contained no detectable levels of lectins. Whilst trace amounts were detected in the embryonic axes of 2 day old seedlings, none of the aerial parts of 3 week old seedlings contained detectable lectins and there have been no reports of these proteins being detected elsewhere in the plant. Thus the

castor bean lectin genes appear to be expressed in an essentially seed-specific fashion. Under the conditions employed by Harley and Beevers, the lectin levels in the endosperm had dropped by 50% on day 4 of germination. This decline was found to continue even more rapidly thereafter with the endosperm tissue eventually being completely consumed.

I.3.B. General properties of castor bean lectins

Lectins may be defined as proteins or glycoproteins which have the ability to bind to carbohydrate species. The binding of lectins to mono- or oligo-saccharides normally occurs with great specificity. Most lectin molecules contain at least two sugar binding sites and therefore are able to agglutinate cells by forming cross-links between the carbohydrate structures present on cell surface glycoproteins or glycolipids. This effect was first reported by Stillmark (1889), who observed the agglutination of blood cells by castor bean extracts. It was subsequently found that two types of lectin are present in the castor bean seed. Originally these lectins were named RCA (Ricinus communis agglutinin) I and RCA II, but more recently RCA II has become commonly known as ricin. For the sake of clarity, the two lectins will be referred to hereafter as RCA I and ricin. The approximate

molecular weights of these lectins are 120 KD and 60 KD respectively.

Lectins have in fact been found in a wide variety of cells, including those of animal and bacterial species, as well as plants. Plant lectins are the most abundant and best studied of these three groups (Liener, 1976). Over 100 different plant lectins have been described, yet the actual biological function of most of these proteins is not clear (Goldstein and Hayes, 1978). This contrasts with other taxonomic groups such as bacteria where examples of assigned functions include the binding of pathogenic species to host cells. One example of a suggested physiological role for a plant lectin is in fact in the binding of nitrogen-fixing Rhizobia to legume roots (Dazzo and Sherwood, 1983, Diaz et al, 1989). Other suggested functions include a possible protective role against fungal pathogens (Mirelman et al, 1975).

Both lectins present in castor bean endosperm tissue are composed of two different types of subunit. These subunits form a heterodimer in the case of ricin and a heterotetramer in the case of RCA I. The ricin subunits are normally known as the A and B chains respectively, whilst the subunits of RCA I are usually referred to as the A' and B' chains. This terminology is appropriate, since the corresponding subunits of the two lectins are

known to be closely related structurally (Nicolson et al, 1974), serologically (Pappenheimer et al, 1974) and at the primary sequence level (Roberts et al, 1985). In the case of ricin, the 32 KD A chain is linked covalently via a single disulphide bond to the 34 KD B chain. RCA I consists of two ricin-like heterodimers held together by noncovalent forces (Olson et al 1974). Again the A' and B' chains (32 KD and 37 KD respectively) are linked by a single disulphide bond. The subunits of ricin and RCA I are all N-glycosylated (Olson and Pihl, 1982), and the difference in molecular weight between the respective lectin B chains is due to differences in their glycosylation patterns. Thus the unglycosylated precursors of ricin and RCA I are of essentially identical molecular weight (Butterworth and Lord 1983). Roberts et al (1985) showed from the nucleotide sequences of ricin and RCA I cDNA clones that the respective lectin A chains share 93% amino acid homology whilst the corresponding B chains are 84% homologous. Thus there is little doubt that the two proteins have evolved from a common ancestor. In view of the structural and sequence similarities described, it is perhaps surprising that ricin and RCA I have strikingly different biological properties. As previously mentioned, RCA I is a potent haemagglutinin, whilst ricin is only weakly so. On the other hand, whilst RCA I is relatively non-toxic to cells, ricin is one of the most potent cytotoxins known to man (Olson and Pihl,

1982). The toxic effects of ricin are in fact due entirely to the A chain, which inhibits cellular protein synthesis by the enzymic inactivation of ribosomes (Endo and Tsurugi, 1987). The B chain nonetheless plays the essential role of delivering the A chain into the cells. This is achieved by the opportunistic binding of the B chain to cell surface glycoproteins and glycolipids (Olanes et al., 1974), followed by internalisation of the A chain. The intoxication mechanism of ricin could be summarised as follows:-

1) Binding to surface of cell

Both ricin and RCA I bind to exposed galactose residues on cell surfaces in a stereospecific and reversible fashion. Studies with RCA I have shown that the oligosaccharide side chains attached to the B chain are not required for its galactose binding activity (Podder et al., 1974). Although it has been estimated that HeLa cells each contain as many as 3×10^9 potential ricin binding sites (Sanvig et al., 1978), the internalisation of A chain molecules is in fact highly inefficient. Thus even though it is thought that cell death may result from the presence of a single ricin A chain molecule in the cytoplasm (Elklid et al. 1980), it appears that several thousand ricin molecules must bind to the cell surface in order to create the statistical probability of a single molecule gaining cytoplasmic entry (Olanes et al., 1979).

ii) Internalisation

After binding of the ricin molecule to the cell surface has taken place, a finite interval occurs before intoxication effects are observed (Refnes et al, 1974). This presumably corresponds to the time taken for the toxin to be transported to the cytosol. Although no defined mechanism for the internalisation of ricin A chain has yet been proposed, most evidence suggests that the toxin molecules are taken in by endocytosis. Studies using electron microscopy have shown that internalised ricin molecules are transported in the cell's Golgi apparatus, endoplasmic reticulum and lysosomes, with at least some of the toxin being transported back to the cell surface (Goyatas et al, 1975, 1980). It seems likely that molecules which evade this pathway are responsible for the eventual intoxication of the cell. Current theories on the above process have recently been extensively reviewed by Olanes and Sandvig (1988).

iii) Inhibition of protein synthesis

Extensive studies on the mode of action of ricin A chain have shown that inhibition of protein synthesis occurs by the catalytic inactivation of the 60S ribosomal subunit (Olanes and Pihl, 1976). Although the 80S ribosomes of both animal and plant cells are sensitive to ricin, the 70S ribosomes of prokaryotes are resistant to the toxin.

Cawley et al (1978) demonstrated that the interaction of ricin A chain with ribosomes followed Michaelis-Menton kinetics. Kinetic experiments showed that one molecule of ricin A chain could inactivate 1500 salt-washed ribosomes per minute. Endo et al (1987) demonstrated that ricin A chain is capable of modifying the 28S RNA of rat liver ribosomes, rendering the RNA susceptible to cleavage at a specific position when incubated with aniline or mild alkali. Although the nature of the enzymic reaction which gives rise to this phenomenon has not yet been clearly established, it was suggested that the ricin A chain molecule may act as a specific N-glycosidase. Such a mechanism involves the removal of a specific adenine molecule, causing weakening of the neighbouring phosphodiester bonds which would then become more sensitive to chemical hydrolysis. The specific adenine residue removed lies in a highly conserved sequence of rRNA which, in prokaryotic 23S rRNA, has been implicated in the binding of EFG (of which EF2 is the eukaryotic equivalent). Although the exact functional significance of this phenomenon is unclear, these observations have provided a useful system by which to assay for ricin A chain intoxication.

The potent toxicity displayed by the ricin molecule has also prompted much research into its potential applications in cancer therapy. By eliminating the

galactose binding sites of the ricin B chain either chemically or by genetic manipulation and then coupling the modified ricin molecule to cancer-specific antibodies, it is hoped that new molecules capable of selectively killing neoplastic cells might be produced. Such molecules are now commonly referred to as immunotoxins (Lord et al, 1985).

I.3.C. Synthesis and processing of castor bean lectins

Ricin and RCA I are simultaneously synthesised in approximately equal amounts in castor bean endosperm cells during and after the period of testa formation. As previously mentioned, the work of Butterworth and Lord (1983) showed that both lectins are synthesised as a 60.5 KD precursor containing the respective A and B chain polypeptides. When mRNA extracted from developing castor bean endosperm is translated in vitro in the presence of dog pancreatic microsomal vesicles, a group of 64-68 KD molecular weight polypeptides immunoprecipitable with antibodies raised against each lectin subunit is produced (Roberts and Lord, 1981B.). It was shown that these polypeptide species had been cotranslationally translocated into the lumen of the vesicles. Furthermore, it was demonstrated that the translocation of the lectin precursors was accompanied by the removal of an N-terminal signal sequence and by the addition of oligosaccharide

side chains to appropriate asparagine residues. The size heterogeneity of the translocated polypeptides was shown to be due to heterogeneity of glycosylation, since a single 58 KD polypeptide was seen after the lectin oligosaccharide chains had been enzymically removed using endo-N-acetylglucosaminidase H. This 58 KD polypeptide corresponded to the two respective lectin precursors lacking both their oligosaccharide chains and their N-terminal signal sequences (Roberts and Lord, 1981B).

The subsequent fate of the glycosylated, N-terminally processed lectin precursor molecules has been examined by means of in vivo radiolabelling (Lord et al, 1984). Fractionation of developing castor bean endosperm tissue previously treated with ^{35}S -methionine has confirmed that a group of glycosylated polypeptides appears initially in the endoplasmic reticulum lumen. When in vivo glycosylation is inhibited by treatment of the tissue with tunicamycin prior to radiolabelling, the 58 KD polypeptide mentioned above can again be observed. By chasing the ^{35}S -methionine initially added with unlabelled methionine, it is possible to follow the subsequent route of the glycosylated lectin precursors. Initially, the labelled polypeptides secreted into the endoplasmic reticulum were observed to disappear, whilst authentic size lectin subunits appeared in a soluble fraction which included components of the protein body matrix. The pathway taken

by the lectin precursors between the endoplasmic reticulum and protein bodies is still uncertain. However, there is evidence to suggest that the Golgi apparatus is in some way involved. Firstly, it has been shown that after leaving the endoplasmic reticulum, the radiolabelled precursors appear briefly in a denser vesicle fraction, where alterations to the oligosaccharide chains occur. This includes the addition of fucose residues, a phenomenon characteristic of the Golgi apparatus (Schachter, 1974). With specific regard to the biosynthesis of ricin, it is important that the toxic A chain does not at any time come into contact with the plant ribosomes. It appears that there is a membrane barrier surrounding the lectin polypeptides at all times during the biosynthetic pathway, thereby preventing any toxin-ribosome interactions. Furthermore, there is no evidence that any of the ricin precursor forms are enzymically active. It has in fact been shown that proricin has no ribosome modifying activity but is able to bind to galactose (Richardson, P. T., Westby, M. J., Roberts, L. M., Gould, J., Colman, A., and Lord, J. M., (1988) Eur J Biochem (submitted)).

I.3.D. Primary structure and folding of castor bean lectins

The complete primary sequence of the ricin A and B chains has been determined both directly by protein sequencing (Funatsu et al., 1980) and indirectly by DNA sequencing of cDNA clones (Lamb et al., 1985). The primary sequence of the RCA B chain has also been determined by direct protein sequencing (Araki et al. 1986) and RCA I cDNA clones have been isolated and sequenced (Roberts et al., 1985). Despite the obvious differences in subunit interactions between ricin and RCA I, the structural arrangement of the two lectin precursors appears to be essentially identical (Roberts et al., 1985).

Both lectin precursors appear to contain a 35 amino acid signal peptide of identical composition. The ricin cDNA clone described by Lamb et al. (1985) has in fact been shown to be incomplete at the 5' end, lacking the methionine ATG start codon (Halling et al. 1985). Since the ricin and RCA I signal peptides have identical primary sequences throughout the region encoded in the cDNA clones, differing at the nucleotide sequence level by only 1 bp, it is assumed that the conclusions of Halling et al. can also be applied to the RCA I gene. This assumption is borne out by the fact that only a single size polypeptide is seen when castor bean endosperm mRNA is translated in vitro and the products immunoprecipitated with either ricin or RCA I specific antibodies (Butterworth and Lord, 1983). Primary sequence data shows that the A and B

chains of ricin consist of 267 and 262 amino acids respectively. Whilst the B chain of RCA I was found to contain the same number of residues as that of ricin, it has been shown that the RCA A chain is one amino acid shorter than its ricin counterpart. This is due to it lacking an alanine residue present at position 130 in the ricin A chain. The two lectin A chains are 93% homologous at the amino acid level, differing in their primary sequence at only 18 positions. The degree of homology between the corresponding B chains is, however, lower (84%), with 41 residues differing.

The A and B chain peptides are joined in both the ricin and RCA I precursor proteins by a linker region containing twelve amino acid residues. The amino acid sequences of the ricin and RCA linkers are identical. These linker regions are cleaved from the respective precursors during the maturation process. In total, ricin and RCA I were found to vary in their primary sequence at 59 positions. Roberts et al., (1985) noted that at 22 of these positions, the respective amino acids of the ricin and RCA I polypeptides were similar in chemical character (as defined by Von Heijne, 1983). It is therefore presumed that the major differences in properties between ricin and RCA I must be determined at locations amongst the remaining 37 points of primary sequence divergence. It was noted that the A chain of RCA I contains four cysteine

residues, only 2 of which are present on the ricin A chain, allowing the possibility for the formation of an extra intra-chain disulphide bond. The B chains of both lectins contained the same 9 cysteine residues, however. This allows the formation of four intra-chain disulphide bonds, plus the inter-chain disulphide bond which ultimately links the mature subunits. There may be some significance in the fact that the ricin and RCA I B chains show different patterns of proline residues in their respective primary sequences. The RCA I B chain has a proline at position 311 which is not present in ricin whereas it lacks the proline present at position 532 on the ricin B chain. Primary sequence variations between ricin and RCA I accounting for their different glycosylation patterns were also observed. Lamb et al (1985) identified four potential sites for asparagine-linked glycosylation in the ricin precursor, at residues 10 and 236 in the A chain and at residue 374 and 414 in the B chain. Roberts et al (1985) showed that the RCA I precursor contained in addition to these four N-glycosylation sites a further site at residue 357 of the B chain.

Whilst it is immediately obvious that the ricin and RCA I genes have arisen from a gene duplication event, it has also been shown that the ricin B chain is the product of a gene duplication (Villafranca and Robertus, 1981). By

aligning the two disulphide bonds on each half of the molecule, 41 homologous pairs of amino acids can be identified. This model is consistent with the observation that the ricin B chain can bind two molecules of galactose at different sites (Houston and Dooley, 1982). The bilobal structure proposed for the ricin B chain is also applicable to the RCA I B chain (Roberts et al., 1985).

I.3.E. Structure and organisation of castor bean lectin genes

I.3.E.1) Structure of mRNAs

Nucleotide sequence analysis of both ricin and RCA I cDNA clones has provided details not only of the lectin primary sequence, but also of the structure of the genes themselves. Roberts et al. (1985) showed that whilst the two genes show divergence at the nucleotide sequence level, there is apparently no difference in their overall structure. The open reading frame of each lectin mRNA contains coding sequence organised so as to produce a precursor with the structure previously described. Thus, the most 5' region of the mRNA open reading frame codes for a signal peptide and precedes a region coding for the A chain polypeptide. Immediately 3' to the A chain coding sequence is a region coding for the linker peptide. Finally the B chain coding sequence occurs at the 3' end of the open reading frame. Although it was originally

thought that both lectin cDNA clones described in the literature contained the respective 5' untranslated regions of the genes, this was subsequently disproved. Halling et al., (1985) showed from the nucleotide sequence of a ricin genomic clone that the ricin signal peptide is most probably 35 amino acids in length rather than 24 amino acids as suggested by Lamb et al., (1985). Thus the extreme 5' end of the ricin and RCA I cDNA clones originally described must contain coding sequence for the respective signal peptides, rather than the 5' untranslated regions of the mRNAs. Halling et al., (1985) showed by S1 nuclease mapping that the 5' untranslated region of the ricin gene contained in their genomic clone was approximately 35 bp in length. Two sizes of S1 nuclease protected band were observed in this experiment, differing in size by 1 bp. Although the doublet of bands observed may be an artefact of the digestion process, it could alternatively indicate that transcription is initiated at two points. The above results were in approximate agreement with primer extension studies carried out by Lamb (1984). Unfortunately there is no information on the RCA I 5' untranslated region as the cDNA clone described by Roberts et al. lacked this section of the mRNA and no genomic sequences have yet been published. The RCA I cDNA clone does, however, contain 69 bp of the 3' untranslated region. The sequence of this portion of the clone shows over 97% homology with that of

the ricin cDNA, but does not extend as far as the polyadenylation site. The ricin cDNA clone described contains a full-length 3'-untranslated region of 159 bp. Halling et al., (1985) suggested that the polyadenylation site of their ricin genomic clone (pAKG) occurred 148 bp downstream from the translation termination point. This conclusion was reached by comparison with a virtually identical cDNA clone (sequence unpublished) containing a complete 3' end. The genomic and cDNA ricin clones described have both been found to contain two polyadenylation signals. One is of the canonical form AATAAA (Messing et al., 1983) and occurs 58 bp downstream from the translation termination point. The other is of the form AATAAG and occurs 20 bp or 25 bp upstream from the polyadenylation site.

By comparing the ricin genomic sequence with that of the cDNAs isolated, Halling et al., (1985) demonstrated that no untranslated intervening sequences were present in their ricin genomic clone. The absence of introns in the ricin gene is not unexpected, in view of the fact that a number of other plant genes share this feature. These include several lectin genes, such as those of soybean (Vodkin et al., 1983), French bean (Hoffman, 1984) and pea (Kaminski et al., 1987), as well as a number of other plant storage protein genes such as the zein genes of maize (Hu et al., 1982).

I.3.E.ii) Genomic flanking sequences of a ricin gene

To date, the only castor bean lectin genomic clone described in the literature is the pAKG ricin gene isolated by Halling et al, (1985). The published sequence extends 257 bp upstream from the probable cap site at the 5' end and 165 bp downstream from the most likely polyadenylation site at the 3' end. As is common with many other plant genes, the ricin gene 5' flanking region was found to be A+T rich. Two putative transcriptional regulatory elements were identified in this region. Firstly a TATAA sequence was located 33 bp upstream from the putative transcription start site. A second sequence reading TATTAA was found 4 bp upstream from the first TATA box, giving an overall sequence reading TATTAATTTTATAA. Halling et al observed that a near perfect 19 bp direct repeat of this element occurs 61 bp further upstream and suggested that it might signify the presence of dual promoters.

Still further upstream, Halling et al identified another element possibly equivalent to the "CAAT" or "AGGA" box found in the 5' flanking regions of other plant genes (Messing et al, 1983). It was suggested that the sequence CAAGT occurring at -93 bp might fulfil the same function as a CAAT box. Immediately 5' to this region, however, is another sequence of possible significance which reads,

together with the first element, ATTAGATCAAGT. The first five bases may possibly be analogous to the "AGGA" box described by Messing et al., (1983), since the trinucleotide sequence (G/T)NG is flanked on either side by an adenine residue. The number of adenine flanking residues normally observed is, however, two to five. Three canonical CAAT sequences were identified at -153 bp, -190 bp and -194 bp. Although these fit the consensus sequence completely, such elements would not normally be expected to occur this far upstream. Halling et al. (1985) did not identify any sequences in the pAKG ricin gene promoter which might be involved in the seed-specific expression of the gene, but Dickinson et al. (1988) have since pointed out that a TATGCATA sequence is present at -255. As will be discussed later, this element closely resembles the consensus sequence of the RY repeat element found in the promoters of a large number of seed storage protein genes (Dickinson et al., 1988).

iii) Genomic organisation of castor bean lectin genes

Halling et al., (1985) showed by Southern blotting and hybridisation that a number of lectin genes occur in the genome of R. communis. It was estimated that this gene family comprises at least six genes. Circumstantial evidence for the existence of a gene family was previously obtained by Cawley et al., (1978), who found that they

could distinguish 3 forms of ricin and 2 forms of RCA I by isoelectric focusing. Although these isoelectric variants might be the polypeptide products of different genes, the possibility remains that these distinct species could have arisen from a single precursor. For instance, the heterogeneity observed might be caused merely by a heterogeneity in the glycosylation pattern of the proteins. Although post-translational modifications may play some part in this phenomenon it has nevertheless been shown that at least 2 forms of ricin with distinct primary sequences exist in some varieties of R. communis. Mise et al. (1977) isolated a variant form of ricin, subsequently referred to as ricin E, from small grain castor bean seeds. Ricin D, the commonly studied form of ricin also occurs in seeds of this type, whereas large grain castor bean seeds were found to contain ricin D only. Araki and Funatsu (1987) demonstrated that the primary sequence of the ricin E B chain differs from that of the ricin D B chain, whereas the respective A chains appeared to be identical in their amino acid composition. In fact, the B chain of ricin E appears to be a hybrid protein, since the primary sequence of its N-terminal half matches that of ricin D, whereas the C-terminal half resembles that of RCA I. Araki and Funatsu therefore proposed that ricin E is the product of a gene which arose by recombination between the ricin D and RCA I genes. A cDNA clone for ricin E was subsequently described by Ladin et al., (1987). The

nucleotide sequence of this clone matched the ricin E primary sequence data obtained by Araki and Funatsu. It is interesting to note however that the sequence of the 3' untranslated region matches that of ricin rather than RCA I at the two residues of divergence. Nothing is known of the ricin E 5' untranslated region, as the cDNA clone described was full-length only at the 3' end.

SECTION I.4. DEVELOPMENTAL AND TISSUE-SPECIFIC EXPRESSION OF PLANT GENES

I.4.A. Analysis of plant gene expression

It appears that in most of the plant systems studied, gene expression is regulated principally at the transcriptional level. The developmental expression of most plant genes may therefore be conveniently monitored in terms of mRNA abundance. It should be noted that changes in the level of a given transcript in vivo may be due not only to changes in the rate of its synthesis, but also to changes in the rate of its degradation. A notable example where the latter predominates is in the cytokinin modulation of rubisco small subunit and cab gene transcript levels in Lemna gibba (Flores and Tobin, 1988). Transcriptional levels may be quantified either by dot blotting (Traina and Cohen, 1983), Northern blotting (Williams and Mason, 1985) or nuclease protection (Weaver and Weissmann, 1979, Kreig and Melton, 1987). The latter two techniques were used in this study and are described fully in the methods section. The importance of transcriptional control in regulating the expression of a given gene may be assessed by monitoring the levels of both the mRNA and protein product of the gene throughout development. A high level of transcriptional control may be inferred if the

developmental profiles of transcript and protein accumulation are similar.

I.4.B. Heterologous gene expression as a tool for analysing cis-acting regulatory sequences

The identification of promoter sequences controlling the tissue-specific and developmental expression patterns of plant genes is not normally possible by DNA sequence analysis alone. Firstly there may be insufficient data available if only one or a few related genes have been cloned. Secondly, although a comparison of the gene flanking sequences may reveal the presence of conserved regions, no functional significance may be inferred from this evidence alone. This is particularly true of genes which have arisen relatively recently from a common ancestor, since a high degree of shared homology would be expected even in non-critical regions. Cis-acting promoter elements must therefore be characterised by functional studies. The most usual approach is to mutate sequences in the promoter region and examine the effects of such mutations on promoter activity. A suitable expression system must therefore be available for the assay of promoter mutants. The various strategies used are summarised below.

1.4.B.1) Heterologous expression in transgenic plants

Transformation systems are now available for a large number of higher plant species. To date, the most commonly used systems have been based on the Ti-plasmid of Agrobacterium tumefaciens. The bacterial pathogen responsible for crown gall disease in a wide range of higher plants (Zambryski et al, 1980). Pathogens of this type are able to introduce their own DNA into the genome of the host plant. The DNA appears to integrate randomly in the host cell genome and with variable copy number. In the Ti plasmid of A. tumefaciens the transferred DNA segment is known as the T-DNA. Non-oncogenic Ti plasmids have been produced by genetic manipulation for the purpose of transferring foreign DNA segments of interest into a wide range of higher plant species. (eg Bevan, 1985, Jones et al, 1985).

If heterologous gene expression is to be used as a means of assaying promoter activity, it is essential that the gene promoter of interest is recognised in the correct way by the cellular machinery of recipient plant species. In general, this appears to be the case. For example, a number of seed protein genes, such as a soy bean beta-conglycinin gene (Beachy et al, 1985) and a pea legumin gene (Ellis et al, 1988) have been expressed in heterologous transgenic plants under the control of their own promoters. It was found in both examples that the

transferred genes were expressed in the correct developmental and tissue-specific fashion. Furthermore, the proteins encoded by these genes were synthesized and deposited in the same fashion as in their native plant species. The ability of plant promoters to function in heterologous systems has been demonstrated with various other types of genes, including light regulated (Schreier et al., 1985), wound inducible (Sanchez-Serrano et al., 1987), heat inducible (Baumann et al., 1987) and nodule specific (Jenson et al., 1986) genes. An exception to this general rule was found, however, when a maize 19 kd zein gene was expressed in Petunia (Ueng et al., 1988). Although transcription was observed to occur at a low level, the heterologous gene was not transcribed in the correct tissue-specific fashion and no protein products could be detected. These results might be attributable to differences between the monocot and dicot gene regulatory systems. It has been shown, however, that the promoter of another maize zein gene is able to direct the endosperm-specific expression of a reporter gene (see below) in transgenic tobacco plants (Scherthauer et al., 1988). Since tobacco and Petunia are close phylogenetic relatives, the results of Ueng et al. (1988) may not be attributable to monocot/dicot promoter incompatibilities.

1.4.B.ii) Transient expression systems

One important disadvantage of using transgenic plants as a

means of assaying promoter activity is the length of time needed for results to be obtained. An alternative to the above procedure is to express heterologous genes transiently in protoplasts. Krens et al (1982) showed that tobacco protoplasts could be transformed by a procedure which involved polyethylene glycol and calcium nitrate treatment. Other techniques used for protoplast transformation include electroporation (Lindsay and Jones, 1987) and particle bombardment (Wang et al, 1988). Lipphardt et al (1988) have successfully used transient expression as a means of identifying the regulatory cis elements involved in the uv induction of an Antirrhinum chalcone synthase gene in parsley protoplasts. This approach may therefore provide a simple and rapid way for promoter activity to be analysed. One obvious drawback is that the physiological conditions in protoplasts differ considerably from those in intact plants. Transient expression may therefore be unsuitable for the analysis of promoters controlling tissue specific genes, such as seed storage protein genes. Nonetheless the example above illustrates the considerable potential of transient expression as a means of analysing inducible promoters, such as those of photosynthetic and heat shock genes.

I.4.B.iii) Reporter genes

The ease with which plant gene promoters or promoter mutants can be analysed in heterologous systems depends to

a large extent on the ease with which gene expression can be monitored. In order to make this procedure as simple and rapid as possible, an assayable "reporter" gene is frequently used. Reporter genes commonly used in plant systems include the E. coli chloramphenicol acetyl transferase (CAT) gene (Jenson et al., 1986), the A. tumefaciens Ti plasmid nopaline synthase (nos) gene and the E. coli transposon Tn5 neomycin phosphotransferase gene (Schreier et al., 1985). More recently, however, the E. coli beta-glucuronidase (GUS) gene has become the preferred choice, as it is very stable and detectable in very small amounts. Since most higher plant species lack intrinsic beta-glucuronidase activity, this assay system is particularly sensitive. Beta-glucuronidase may be detected fluorometrically from very small amounts of plant material or in situ by light microscopy after the tissue has been treated with an appropriate chromogenic substrate.

I.4.C. DNA binding proteins

Although very little is known about the transcriptional regulatory mechanisms of higher plants, it is clear that a number of factors in the nucleus are involved in the determination of promoter specificity. The RNA polymerase II molecule of eukaryotic cells differs from its prokaryotic counterpart in that it lacks the inherent

ability to recognise promoter sequences in vitro. Some of the nuclear factors required for RNA polymerase II transcription in animal systems appear to be ubiquitous in their occurrence (Dylan and Tjian, 1985). Other factors, however, have been shown to be promoter specific. It is presumed that the latter group are involved in developmental and tissue specific patterns of gene expression. Protein species which bind specifically to the upstream regions of animal genes (Dylan and Tjian, 1985) and plant genes (eg Maier et al, 1987, Graen et al, 1987, Deikman and Fischer, 1988) have been identified and, in some cases, characterised. Unfortunately, far less is known about the DNA binding proteins of plant systems than their animal counterparts. DNA binding proteins are often known alternatively as trans-acting factors on account of their mode of action and are commonly detected either by gel retardation (Fried and Crothers, 1981, DNase footprinting (Galas and Schmits, 1978) or filter binding assays (Diffley and Stillman, 1986). The first eukaryotic DNA binding protein to be characterised was the sp1 protein required for the transcription of SV40 viral genes in cultured animal cells. It was shown that sp1 binds to a well defined upstream element occurring 50-100 bp upstream from the SV40 early viral transcription start site, thereby activating transcription (Fromm and Berg, 1982). The sp1 binding element appears to function bidirectionally (Everett et al, 1983). Sp1 binding sites

have in fact been detected in the promoter regions of many animal and animal viral genes. This raises doubts as to the promoter specificity of *sp1*. It has been suggested that *sp1* might provide a structural framework to which other regulatory proteins might bind, or a basal, constitutive level of transcription which may be modulated by other factors (Dylan and Tjian, 1985). The binding of nuclear proteins to promoter sequences in animal systems have been correlated with a number of different effects, including the switching on and off of transcription, the modulation of transcriptional levels and tissue specific transcription. These phenomena have been extensively reviewed (Khoury and Gruss, 1983, Dylan and Tjian, 1985, Joss et al, 1986).

The most extensively characterised plant DNA binding protein is the GT-1 protein which appears to be involved in the regulation of rubisco small subunit gene expression in pea (Green et al, 1987). The possible functional role of this protein is discussed below.

I.4.D. Identification of cis-acting DNA sequences

As previously mentioned, the identification of cis-acting regulatory sequences is normally achieved by examining the effects of mutation on the functional activity of the promoter in question. Exonuclease digestion is commonly

used as a means of creating promoter mutants, since it allows the rapid production of a sequential series of deletions. The same deletions used for functional analyses of this type may also be assayed for their ability to interact with DNA binding proteins. Since the functional role of trans-acting DNA binding proteins is intimately linked with that of the cis-acting DNA elements which they recognise, these two aspects will be discussed together.

I.4.D.1) Photosynthetic genes

The two most extensively studied photosynthetic genes are those which encode the small subunit of rubisco and the chlorophyll *a/b* binding (*cab*) proteins respectively. Both genes are light inducible and appear to occur in most plants as a multigene family. Simpson et al (1986) showed that a 247 bp (-100 to -347) fragment from a pea *cab* gene promoter was capable of conferring light inducibility on a normally constitutive nopaline synthase (nos) promoter driving the expression of an NPTII gene expressed in transgenic tobacco plants. In light conditions, transcription of the NPTII gene was enhanced. Moreover, this effect was shown to be independent of the orientation of the fragment. The *cab* promoter fragment could therefore be said to be acting as a tissue specific enhancer. Simson et al showed further that the same DNA fragment behaved as a tissue specific "silencer", since it

repressed nos promoter-driven expression of the reporter gene in the non-photosynthetic, such as the roots of the same transgenic tobacco plants. Ha and An (1988) showed by promoter deletion analysis that at least three cis-acting elements are involved in the light-inducible expression of a cab gene from Arabidopsis thaliana. One of these elements, which was located between -158 and -253, was sufficient to confer promoter specificity, but another element located between -253 and -321 was required for maximal gene expression. A further element located between -766 and 1396 appeared to act as a promoter specific modulator. Ha and An showed by further manipulations that the position of an 8 bp Z-DNA -like sequence correlated well with the activity of the first element described. Hence this potential Z-DNA sequence may well be important for cab promoter specificity. The presence of potential Z-DNA forming sequences in a number of plant gene promoters had been noted previously (An, 1987). Z-DNA sequences have in fact been found in nearly all of the plant promoters investigated to date, suggesting that Z-DNA structures may play an important role in plant gene expression. It is interesting to note that in the Arabidopsis cab gene promoter described previously, there are two additional Z-DNA forming sequences which are inverted repeats of the first element. The position of these elements appears to correlate with

the position of the two other regulatory elements postulated from deletion studies.

The gene encoding the small subunit of rubisco (rbcs) shows a light-inducible and tissue specific expression pattern similar to that of the cab gene. Morelli et al (1985) showed by deletion analysis that a 33 bp sequence close to the TATA box of the pea rbcs-E9 gene promoter is sufficient to confer light-inducibility. This sequence is highly conserved amongst higher plant rbcs genes. It was shown that for maximal transcription of the rbcs gene, an upstream region of approximately 700 bp is required. A more detailed analysis of the cis-acting elements in the pea rbcs-3A gene promoter was described by Fluhr et al (1986). The functional data obtained was correlated with sequence conservations observed between the rbcs-3A and rbcs-E9 genes. It was noted that a conserved enhancer-like sequence was present not only in the 33 bp region surrounding the TATA box, but also further upstream at approximately -150. The functional data obtained suggested that sequences in the -150 region show both enhancer and photoregulatory properties, as had previously been observed with the sequence surrounding the TATA box. It therefore appears that former distinctions made between enhancer elements and tissue/developmental specific elements might be an oversimplification. Three conserved boxes (designated I, II and III), which are conserved

amongst all known pea rbcS genes, were identified in the upstream enhancer region. Fluhr et al termed these boxes light responsive elements (LREs). Two boxes homologous to boxes II and III (designated II* and III* respectively) were also identified around -220. The latter two boxes are thought to be redundant, but function in vivo when boxes I, II and III are deleted (Kuhlemeier et al, 1987). Green et al (1987) identified a protein factor that specifically interacts with both boxes II and III and boxes II* and III*. This protein, known as GT-1, is present in the nuclei of both light and dark grown pea leaves. The involvement of this factor in photoregulation does not, therefore, seem to depend on its de novo synthesis. The binding of GT-1 to its receptor sites correlates well with transcriptional activity, however, so it is possible that the photoregulatory process might involve the post-translational modification of the GT-1 protein or selective blocking of the binding sites by other factors. The identification of the exact functional role of GT-1 is further complicated by results which demonstrate that the protein interacts with sequences in the promoter region of a mitochondrial ATPase B subunit (Boutry and Chua, 1985) and a cauliflower mosaic virus 35S promoter (J F Nagy and P J Green, unpublished). Both the aforementioned genes appear to be expressed in an essentially constitutive fashion, thus casting doubt as to whether GT-1 is specifically involved in photoregulation.

Green et al (1988) suggested that the GT-1 binding sites in these two constitutive genes may contain LREs that are masked by other elements. Alternatively GT-1 may be a general transcription factor involved in the activation of a large number of genes.

I.4.D.ii) Seed protein genes

The promoters of a number of plant seed protein genes have been functionally analysed by deletion studies (Chen et al, 1986, Colot et al, 1987, Roussell et al, 1988) and DNA binding proteins have been identified in some cases (Jofuku et al, 1987, Maier et al, 1987). The zein genes of maize are probably the best studied group of plant seed protein genes. Zein proteins fall into two size classes, of 19 Kd and 21 Kd, and accumulate specifically in the endosperm of maize seeds. Roughly 120 zein genes are thought to occur in the maize haploid genome (Viotti et al, 1979). Roussell et al (1988) showed that an upstream sequence between -337 and -125 is required for the maximal transcriptional activity of a 19 Kd zein gene promoter driving the expression of a CAT gene in carrot protoplasts or sunflower tumour cells. Only very low levels of transcription were observed with constructs containing a 125 bp fragment immediately upstream from the zein gene cap site. Both the CAAT and TATA boxes are present in this fragment. Examination of the 5' sequences upstream

from the -125 position revealed the presence of five regions sharing homology with the SV40 enhancer core sequence, at -321, -301, 271, -232 and -217. Brown et al (1986) carried out a comparison of the 5' flanking sequences of various zain genes encoding the two size classes of zain polypeptide. The largest single region of homology was identified as a 15 bp CACATGTGTAAAGGT sequence at -330. This region contains part of an animal core enhancer sequence and overlaps a 30 bp sequence capable of forming a stem-loop structure. Furthermore, elements of this type have been found at similar locations in a maize sucrose synthase gene (Warr et al, 1985) and a number of cereal storage protein genes (Forde et al, 1985). All the genes described are expressed in an endosperm specific fashion, suggesting that the 15 bp element may be involved in conferring tissue specificity on the promoters. Maier et al (1987) investigated the sequence specificity of a DNA binding protein present in maize seed nuclear extracts which interacts with a zain gene promoter. By using synthetic oligonucleotides, a 22 nucleotide binding site between -318 and -339 was identified. This region contains 14 of 15 bp of the conserved zain gene box. The binding of the nuclear protein factor did not require the presence of the potential stem-loop sequence, suggesting that the interaction depends on sequence specific, rather than three dimensional recognition. It is interesting to note,

however, that the binding site identified contains a potential Z-DNA forming sequence (An, 1987). A second binding site in the region between positions -8 and -255 was also identified, although its position was not determined.

Deletion studies carried out by Colot et al (1987) on a wheat glutenin gene promoter add weight to the hypothesis that a conserved TGTAAAG sequence is in some way involved in the endosperm specific expression of cereal seed protein genes. The fact that the promoters of various cereal endosperm protein genes are correctly recognised in transgenic tobacco plants (Colot et al, 1987, Hoffman et al, 1987, Scherthaner et al, 1988) suggests that common nuclear factors may be present in both. Certain sequence homologies have in fact been observed to occur between the promoters of seed storage protein genes in monocotyledonous and dicotyledonous plants. The most widespread element identified is the RY repeat of consensus sequence CATGCATG (Dickinson et al, 1988). Elements of this type have been identified in the promoters of a number of dicotyledonous seed protein genes, including various pea legumin genes (Lycett et al, 1985), two Phaseolus lectin genes (Hoffman et al, 1985), various Arabidopsis 2S albumin genes (Krebbbers et al, 1988) and the pAKG ricin gene (Halling et al, 1985). RY repeats also occur in many cereal seed protein genes at or

near the -300 element. The RY repeat does in fact occur within a larger region of conserved residues in a number of cases (eg Lycett et al, 1985). The exact functional significance of this sequence conservation clearly requires further investigation.

I.4.D.iv) Future prospects

Although current research on plant promoters lags behind parallel studies in animal systems, the identification of cis-acting DNA elements and their trans-acting DNA binding proteins has been achieved in a number of cases in higher plant systems. The previous distinction made between enhancer and tissue specific elements now appears to be less applicable and there must clearly be more functional studies carried out to identify the role of these factors more precisely. Although Green et al (1988) have correlated the binding of the GT-1 protein to the rbcS gene promoter with transcriptional activity, there may well be other examples of trans-acting factors showing a different mode of action. The elucidation of the functional role of DNA binding proteins should be aided by the isolation of cDNA clones for their genes.

SECTION I.5. MOLECULAR CLONING AND ITS APPLICATION TO PLANT SYSTEMS

I.5.A. Principles and strategies

An essential prerequisite for many lines of investigation into the mechanisms of gene function is the availability of cloned DNA(s) containing the gene(s) of interest. Once a recombinant DNA clone of the relevant gene has been isolated, many types of experiment become possible. Notable examples include:-

- i) The use of DNA sequencing techniques to obtain details of gene structure and primary sequence.
- ii) The use of hybridisation techniques such as Northern blotting, nuclease protection and in situ hybridisation to investigate patterns and levels of transcription in vivo.
- iii) The use of techniques such as "chromosome walking" (Bender et al, 1983) to map the position of the gene of interest within the genome of the organism.
- iv) The transfer of the cloned gene into heterologous organisms. Heterologous gene expression is often used as a means of producing proteins of high value on a large

scale. Furthermore this approach may also be used to produce transformants with novel and possibly economically useful properties.

Broadly, molecular cloning may be defined as the process by which specific DNA sequences are isolated and maintained in a recoverable form within a replicating DNA molecule constructed in vitro. The bacterium Escherichia coli is normally used for the maintenance and propagation of recombinant DNA molecules since it is easily manipulated and its genetics are well characterised. Whilst the various cloning vectors used with E. coli differ in certain respects, there are nevertheless two features which are essential to cloning vectors of all types:-

- 1) A replication origin is required so as to allow the propagation of the molecule in the host cell.
- ii) Suitable restriction endonuclease sites must be present so that the foreign DNA may be inserted and excised from the vector molecule.

A third feature common to plasmid and cosmid vectors is the provision of a selectable marker gene, such as an antibiotic resistance gene. This ensures that cells which lose the recombinant DNA molecule will not be viable when

grown in the appropriate selective medium. Cells containing cloned DNA sequences normally grow less vigorously than their wild-type counterparts in the absence of selection.

The four main types of cloning vector used with E. coli are as follows:-

i) Plasmids

These are extrachromosomal double stranded DNA molecules which are found in both prokaryotic and eukaryotic cells. Most plasmid DNA molecules are circular. Plasmids vary considerably in size, but for the purposes of genetic manipulation, smaller plasmids are easier to work with. Thus many plasmids commonly used for cloning have been obtained by the deletion of large portions of non-essential DNA from their naturally occurring counterparts. One such example is the E. coli plasmid pBR322 (Bolivar et al., 1977), from which many other laboratory plasmids have been derived. Most plasmids are relatively small compared with other types of vector and therefore are easier to manipulate in vitro. A second advantage of plasmids is the relative ease with which DNA can be isolated from recombinants. There are, however, two important disadvantages. Firstly, plasmids are generally not suited for cloning larger size DNA fragments (10 Kb and above). Secondly, the screening of large numbers of plasmid

recombinants is difficult and time consuming compared with other types of vector, particularly those based on bacteriophage lambda.

ii) Bacteriophage lambda vectors

Lambda is a virus or bacteriophage which infects E. coli. The bacteriophage lambda genome consists of a double stranded linear DNA molecule of approximately 50 Kb. The 5' and 3' ends of the genome have single stranded projections of 12 bp which are complementary, thus giving rise to so-called cohesive or cos ends. This allows the DNA molecules to form circular or concatenated structures. Upon entry into the host cell the bacteriophage does not always cause lysis. Instead the viral genome may integrate into the host chromosomal DNA to produce a latent prophage or provirus. This is a phenomenon commonly known as lysogeny. The transition from lysogeny to virulence is dependent of the CI gene of the phage which, in the lysogenic state, encodes a repressor protein which binds to the promoter region of the virulence genes, thereby preventing their transcription. Lysis is normally observed as either a reduction in turbidity and eventual clearing of a liquid culture, or on agar plates as cleared regions (plaques) in a bacterial lawn. Many individual plaques can be grown in a single petri dish and the plating out process is simple and rapid. Thus bacteriophage lambda vectors are especially suited to

applications where large numbers of clones need to be screened. Furthermore, most lambda vectors are able to accommodate large DNA inserts. This is particularly true of the lambda replacement vectors. These vectors contain a non-essential "stuffer" fragment between the two essential regions known as the lambda arms. Once the stuffer fragment has been removed by restriction endonuclease digestion and size fractionation, inserts of typically 8-22 Kb can be accommodated. The upper and lower limits of insert size are dictated by the corresponding range of genome sizes which can be packaged into infective viral particles in vitro. A recombinant DNA molecule must normally be between 78 % and 105 % of the wild type lambda genome size in order to be packaged into a viable phage particle. Some bacteriophage lambda vectors do not require the removal of a "stuffer" fragment prior to cloning and are known as insertion vectors. Although these cloning vehicles cannot accommodate such large DNA fragments as the replacement vectors, all the other advantages of bacteriophage lambda-based cloning are retained. Thus lambda insertion vectors have been found particularly useful for the construction of large cDNA libraries where the clones of interest are derived from low abundance mRNAs.

iii) Cosmids

Cosmids are cloning vectors constructed artificially so as

to combine some of the advantages of plasmid vectors and lambda vectors. Thus they contain essential DNA sequences derived from plasmids and, as the name might imply, the cohesive or cos ends of the lambda DNA molecule. Since very little lambda DNA is present in these vectors, very large inserts (up to about 52 Kb) of foreign DNA can be accommodated. The recombinant DNA can be packaged in vitro into bacteriophage lambda particles which may then be introduced into E. coli by transduction. Once inside the bacterial cell, however, the cosmid replicates as a plasmid and recombinants are therefore screened by colony hybridisation. Cosmids are normally used only for the cloning of genomic DNA. They are particularly suited to this purpose since the large insert sizes make it possible for clusters of genes to be cloned on a single DNA fragment. The various drawbacks of cosmid vectors include the fact that colony hybridisation, rather than plaque hybridisation must be used and that cosmid systems are not as efficient as bacteriophage lambda systems in the recovery of recombinant DNA molecules.

iv) M13 vectors

M13 is a filamentous E. coli bacteriophage containing a single stranded DNA genome approximately 6.4 Kb in size. Vectors based on M13 are used solely for the production of single stranded DNA molecules. The most common use for these molecules is in the dideoxynucleotide sequencing

method of Sanger et al (1977). Single stranded M13 DNAs may also be used for S1 nuclease mapping (Weaver and Weissmann, 1979) and as strand-specific probes for Northern hybridisation (Hu and Messing, 1982). DNA manipulations with M13 vectors are carried out on the double stranded replicative form (RF) DNA in the usual way. By altering the conditions of culture and extraction, either type of DNA may be isolated from any M13 clone.

I.5.B. cDNA cloning

Complementary DNA or cDNA is the term used to describe molecules which have been obtained by the reverse transcription of RNA species. By using DNA polymerase I in tandem with reverse transcriptase, it is possible to generate double stranded cDNA molecules which may be cloned into a suitable vector by various means (Efstradiatis et al, 1976, Roychoudhury et al, 1976, Oyayama and Berg, 1981, Huynh et al, 1984). One of the principal merits of cDNA cloning is that this technique offers the possibility of enriching the resultant library for clones of the desired type. Since most genes are expressed differentially between different tissues and developmental stages, the poly(A)+ RNA used for cDNA synthesis is normally extracted from tissues expressing the gene(s) of interest at the highest possible level.

Thus cDNA clones for abundant mRNAs may be obtained from screening relatively few recombinants. Normally genomic clones of a given gene are isolated using a homologous cDNA clone as a probe. The isolation of cDNA clones is, however, more complicated. If a homologous clone is not available for use as a probe, cDNA libraries are normally screened either with antibodies in the case of expression libraries (Young and Davis, 1983) or with synthetic oligonucleotides homologous to the gene of interest (eg Lamb et al, 1985).

I.5.C. Nucleotide sequence analysis

The two techniques most widely used for the determination of DNA sequences are as follows:-

1) Chain termination method

This technique, described by Sanger et al (1977), is based upon the principle that certain deoxynucleotide analogues can be incorporated into an elongating DNA strand where they prevent further elongation. The experimental approach used is summarised in Figure 2. The exact methodology of this technique is detailed in section II.11.

11) Chemical cleavage method

This technique was first described by Maxam and Gilbert

Figure 2

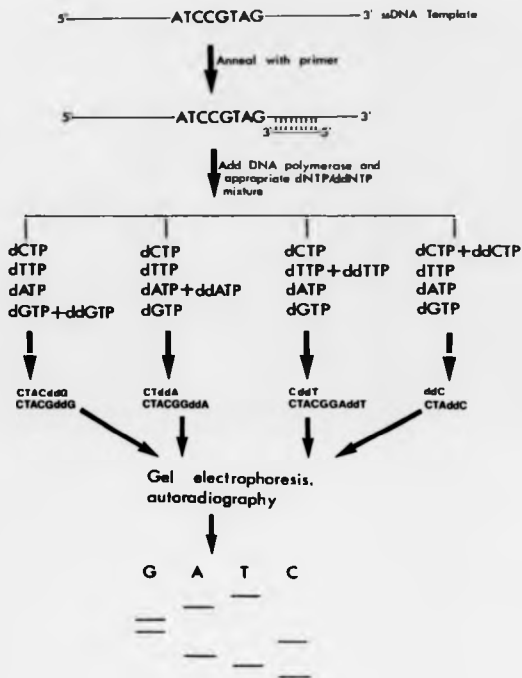


Figure 2 Principle of the chain termination sequencing
method

Many variations on the chain termination sequencing method of Sanger et al (1977) exist, but the rationale is the same for each. The diagram opposite illustrates the main steps involved. A synthetic oligonucleotide molecule is annealed to the single stranded DNA template and serves as a primer for the second strand synthesis reaction, which is catalysed by a DNA polymerase, such as the DNA polymerase I Klenow fragment. Four separate reactions are carried out, each containing a different dideoxynucleotide (ddGTP, ddATP, ddTTP or ddCTP), in addition to the four deoxynucleotides. Dideoxynucleotides can be inserted into an elongating DNA strand at the same positions as their deoxynucleotide counterparts, but further elongation is prevented once a dideoxynucleotide has been incorporated. Thus in each reaction, chain termination will take place at positions corresponding to only one of the four nucleotide types. The newly synthesised DNA strands are separated from the template molecules by heat denaturation and the products of the sequencing reactions are electrophoresed alongside each other on a polyacrylamide gel which is subsequently autoradiographed. As is shown, the nucleotide sequence of the newly synthesised DNA strand can be read from the ladder of bands seen on the gel autoradiograph.

(1976). Figure 3 summarises the experimental procedure involved.

I.5.D. Genomic cloning

I.5.D.1) Choice of vector type

Since most plant genomes are large with respect to their prokaryotic counterparts, large numbers of recombinant genomic clones must be screened in order to isolate a unique sequence of interest. Any vector used for the construction of plant genomic libraries must therefore be suited to large scale screening. A second important criterion is that the cloning procedure should be as efficient as possible, so that a "complete" library may be prepared with a minimum of effort.

Bacteriophage lambda vectors are ideal for the construction of plant genomic libraries as they satisfy both of these criteria. Firstly, the screening of bacteriophage lambda clones by plaque hybridisation is simple and rapid. Secondly, the lambda in vitro packaging reaction gives high yields of recombinant bacteriophages. For the reasons above, it was decided that a bacteriophage lambda vector would be used for the construction of E. communis genomic libraries.

Figure 3

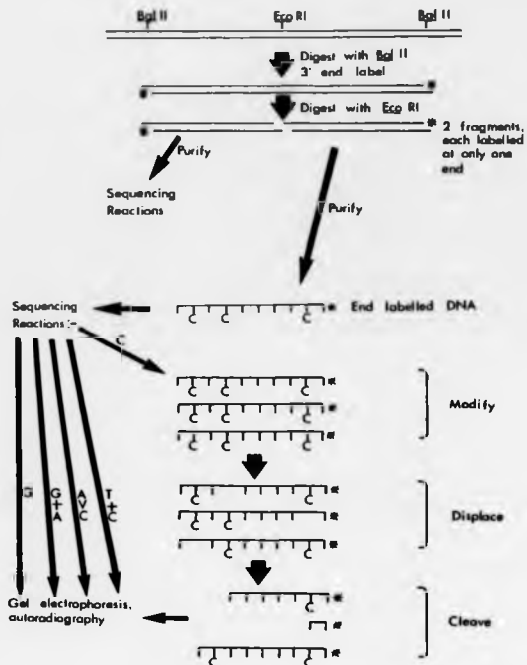


Figure 3 **Principals of the Maxam and Gilbert sequencing method**

An illustration of the basic procedure for sequencing DNA by the chemical cleavage method of Maxam and Gilbert (1976) is shown. Initially, a DNA restriction fragment of interest is excised from the appropriate plasmid. Both strands of the fragment are then radiolabelled at one end only. If the DNA fragment is first 3'-dephosphorylated, the enzyme polynucleotide kinase may be used to add ^{32}P -radiolabelled phosphate groups to the 3' ends of the DNA strand. Alternatively, the strands may be radiolabelled at certain 3' ends using the Klenow fragment of DNA polymerase I. Two radiolabelled DNA strands of unequal size may subsequently be produced by cleaving the restriction fragment with a second restriction enzyme. It is not necessary to purify the radiolabelled DNA strands from their complementary strands, but the two unequally sized double stranded fragments must be purified from each other using polyacrylamide gel electrophoresis. Once purified, the fragments are subjected to a set of different chemical treatments which ultimately result in the base-specific cleavage of the DNA. The cleavage occurs randomly at any one of the appropriate bases in the chain, but the reaction conditions must be optimised so that each DNA strand is only partially cleaved. Table 1 details the chemical conditions used to bring about the base-specific cleavage reactions. It should be noted that many workers omit the A+C reaction, since it is possible to determine the sequence of a DNA strand using the other four reactions alone. Overall, the process may be subdivided into three stages. Initially a set of base-specific modification reactions involving a number of different reagents is carried out. Next, the chemically modified bases are displaced from the DNA strand by piperidine treatment or acid-catalysed depurination. This results in the hydrolysis of the DNA strand at the positions where modification originally took place. Again the sequencing samples are electrophoresed alongside each other on a denaturing polyacrylamide gel which is subsequently autoradiographed. As with the chain termination sequencing method, a ladder of bands is observed from which the DNA sequence may be read.

Table 1

Base-specific Reaction	Base Modification	Displacement	Cleavage
G	dimethylsulphate	piperidine	piperidine
G + A	formic acid	acid-catalysed depurination	piperidine
A > C	NaOH	piperidine	piperidine
T + C	hydrazine	piperidine	piperidine
C	hydrazine/NaCl	piperidine	piperidine

Table 1 Chemical treatments used for Maxam and Gilbert
sequencing

In the DNA sequencing method of Maxam and Gilbert (1976), base-specific DNA cleavage is achieved using a variety of chemical treatments. The sequencing process may be subdivided into three stages; modification, displacement and cleavage, as shown opposite. The exact chemical treatments used are detailed. It should be noted that the A>C reaction is frequently omitted, as DNA sequences may be determined using the other four reactions alone. An account of the Maxam and Gilbert sequencing method is given in Figure 3.

I.5.D.ii) General principles of genomic cloning using
bacteriophage lambda based vectors

In general, lambda replacement vectors are more suitable than lambda insertion vectors for the construction of genomic libraries as they are able to accept larger DNA inserts. Although when using replacement vectors it is first necessary to purify the lambda arms, this is relatively easily achieved. A typical scheme for the construction of a genomic library using a lambda replacement vector is shown in Figure 4.

Three main criteria contribute to the usefulness of a genomic library. These are as follows:-

a) The total number of recombinant clones in the library. Obviously the probability of isolating a DNA sequence of interest will increase with the number of recombinants screened. Clarke and Carbon (1976) showed that the probability of isolating a unique sequence of interest may be calculated using the following formula:-

$$N = \frac{\ln(1-P)}{\ln(1-F)}$$

where N is the number of recombinants in the library

P is the probability of isolating the sequence of
interest

Figure 4

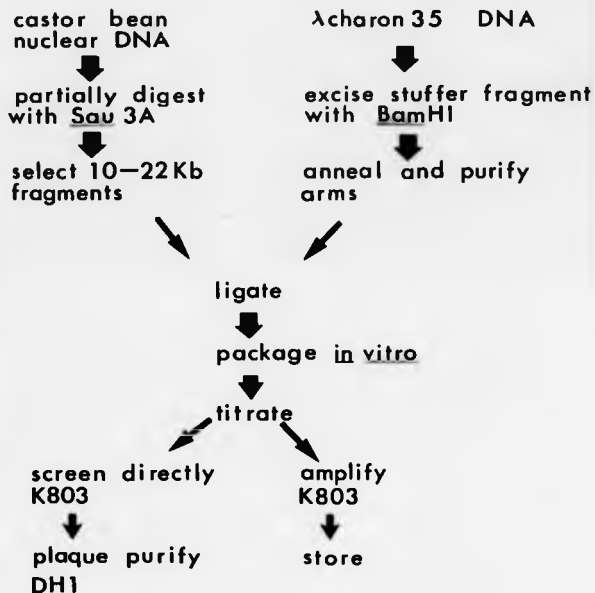


Figure 4 Genomic cloning strategies: Construction
of a castor bean genomic library

An illustration of the procedure used in this study to construct a castor bean genomic library is shown.

F is the fractional proportion of the genome in a single recombinant (ie the average insert size divided by the genome size)

By convention, a probability value of 99% is normally used when calculating the number of recombinants which must be screened. The figure obtained in this way is frequently referred to as the size of a "complete" library.

Obviously the size of this "complete" library depends on two parameters: the haploid genome size (or C-value) of the organism and the average size of the DNA inserts.

One important limitation of the Clarke and Carbon formula is that it relies on the assumption that the population of DNA fragments contained in the library is completely random. The substrate genomic DNA must therefore be fragmented using a method which shows no sequence specificity if the equation is to be applicable. Random fragmentation is usually achieved by subjecting the DNA to mechanical shearing. Maniatis et al (1978) used such an approach to construct genomic libraries from Drosophila and silkworm DNA. The principal merit of this approach is that provided sufficient recombinants are available, it should be possible to isolate any DNA sequence from the genomic library irrespective of the pattern of restriction sites in and around the stretch of DNA in question. This is particularly important when little or no restriction

mapping information is available, or when a library is to be screened for a number of unrelated genes. The use of mechanical shearing as a means of generating genomic fragments of cloneable size is, however, time consuming and rather wasteful of DNA. A far more straightforward approach is to simply digest the substrate DNA with a restriction endonuclease which generates cohesive termini. Obviously the fragments generated in this way will not be truly random (Seed *et al*, 1982). However, the degree of randomness may be maximised in two ways. Firstly, if a restriction enzyme with a 4 bp recognition sequence is used rather than a restriction enzyme recognising a hexanucleotide sequence, the products of a partial digestion will show far greater heterogeneity. Hexanucleotide recognition sites would be statistically expected every 4.1 Kb in random DNA whereas tetranucleotide recognition sequences would be expected to occur every 256 bp. Since the latter distance is small with respect to the average size of the DNA inserts in most genomic libraries, the population of cloneable DNA fragments generated may be considered to be essentially random (Maniatis *et al*, 1982). The second limitation associated with using partial restriction digests to prepare genomic libraries is that the restriction sites recognised by a given enzyme in a genome normally show different susceptibilities to cleavage by the enzyme. This problem may be minimised by digesting separate

samples of the substrate DNA to different degrees. The samples may then be pooled prior to use in cloning. Alternatively a separate library may be prepared from each digest. The task of preparing a genomic library is considerably simplified by using restriction enzymes as a means of generating DNA fragments for cloning. Hence this approach is now the most widely used. The castor bean genomic libraries constructed for the purpose of this study were prepared in this way.

I.5.D.iii) Construction of bacteriophage lambda based genomic libraries

A complete technical description of genomic library construction would be inappropriate here, since the subject is covered in detail in section II.5. The experimental approach used is summarised in Figure 4. By way of an example, the construction of one of the genomic libraries used this study is described. The bacteriophage lambda vector charon 35 (Loenen and Blattner, 1983) was used for this purpose. The essential features of this vector are discussed in section III.2.A.

SECTION I.6. AIMS AND APPROACHES

The broad aim of this study was to investigate the process of seed-specific gene expression in R. communis by means of cloning techniques. More specifically, the objectives were:-

- A. To construct a R. communis genomic library and isolate as many members of the lectin gene family as possible.
- B. To characterise the cloned genes by restriction and nucleotide sequence analysis.
- C. To determine which of the clones genes are expressed in vivo and assess the relative expression levels of the individual members of the gene family.
- D. To characterise at the transcriptional level the in vivo developmental and tissue-specific patterns of the expression of the genes.
- E. To identify promoter sequences which might determine the expression patterns of the lectin genes.

SECTION II

MATERIALS AND METHODS

SECTION II.1. MATERIALS

II.A. Plant material

Ricinus communis seeds of Sudanese origin were kindly provided by Croda Premier Oils, Hull.

II.1.B. Chemicals, biochemicals, radiochemicals and enzymes

The materials used were normally of the highest analytical grade available. The source of specific reagents is given below.

Amersham International PLC, Amersham, Buckinghamshire: [α - ^{32}P]dCTP ($>3000\text{Ci}/\text{mmol}$), [α - ^{32}P]dGTP ($>3000\text{Ci}/\text{mmol}$), [α - ^{35}S]dATP ($>400\text{Ci}/\text{mmol}$), [α - ^{32}P]UTP ($>800\text{Ci}/\text{mmol}$), Hybond-N, DNA polymerase I (Kornberg), DNA polymerase I (Klenow fragment), all restriction endonucleases except Sph I, T4 DNA ligase, T7 RNA polymerase, M13 sequencing kit, bacteriophage lambda CIIndts857Sam7 DNA.

The Boehringer Corporation (London) Ltd, Lewes, East Sussex: calf intestinal phosphatase, ribonucleotides, deoxyribonucleotides, proteinase K, RNAse-free DNAse.

BDH Chemicals Ltd, Atherstone, Warwickshire: acids, Amberlite monobed resin MB-3, ammonium persulphate, bromophenol blue, 2-mercaptoethanol, PEG 6000, phenol, TEMED, xylene cyanol, HEPES, guanidinium chloride.

Calbiochem Corporation, La Jolla, California 92037, USA: aquacide

Difco Laboratories, Basingstoke, Hampshire: bacto-agar, bacto-tryptone, yeast extract.

Eastman Kodak, Rochester, New York, USA: NN'-methylenebisacrylamide.

Fisons PLC, Loughborough, Leicestershire: acrylamide, CaCl₂, organic solvents, SDS.

Gibco-BRL, Paisley, Renfrewshire, Scotland: nuclease-free bovine serum albumin, low melting point agarose (ultra-pure).

Northumbria Biologicals Ltd, Cramlington, Northumbria: restriction endonuclease Sph I

Pharmacia (Great Britain) Ltd, Central Milton Keynes, Buckinghamshire: Dextran T40, Percoll.

Promega Biotech, 2800 S. Fish Hatchery Road, Madison, USA:
Bacteriophage lambda in vitro packaging extracts, human
placental RNase inhibitor.

Schleicher and Schull, Dassel, West Germany:
nitrocellulose type BA85/1

Sigma Chemical Co. Ltd, Poole, Dorset: agarose medium EEO
(type II), ampicillin (sodium salt), ATP, bovine serum
albumin, DNase I (from bovine pancreas), dithiothreitol,
ethidium bromide, IPTG, N-lauroylsarcosine (sodium salt),
polyvinylpyrrolidone-40, RNase A (from bovine pancreas),
spermidine 3-HCl, spermine 3-HCl, E. coli tRNA, gelatin,
BCIG (X-gal), Ficoll 400.

All other chemicals were obtained from either BDH
Chemicals Ltd or Fisons PLC depending on availability.

SECTION II.2. GROWTH AND STORAGE OF BIOLOGICAL MATERIALS

II.2.A. Growth of plantsII.2.A.i) Production of seedlings for the extraction of nuclei

Castor beans were imbibed overnight in running tap water. They were then planted in seed trays at a depth of 1 cm and a density of approximately 1 seed per 2 cm² in pre-moistened vermiculite. The trays were kept in an artificial growth room maintained at 25°C under fluorescent lights providing a 12 h photoperiod. Harvesting normally took place 9 days after planting.

II.2.A.ii) Growth of mature fertile plants for harvesting of seed tissues

Seedlings obtained by the procedure above were transplanted into 5 inch plastic pots containing pre-moistened peat-based compost. The plants were grown in a greenhouse supplied with heating during the winter months to achieve a minimum temperature of 10°C. A 12 h photoperiod provided by tungsten lighting was also maintained during these months. Flowering normally took place approximately 6 months after planting, the period of fruit production lasting for typically 3 months.

II.2.B. Growth and storage of bacterial stocks

The bacterial strains used for this study are listed in Table 2. Storage was carried out as described by Maniatis et al (1982).

For long term storage, bacteria were maintained as frozen glycerol stocks. 850 ul of an overnight culture grown in L-broth (1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 0.2 M NaCl) was mixed with 150 ul of glycerol and stored frozen at -80°C.

Bacterial strains were maintained in the short term on L-agar plates (L-broth solidified with 1.5% (w/v) bacto-agar). A loopful of a fresh overnight culture was spread onto a dried plate and incubated at 37°C overnight. The plates were then stored at 4°C for up to 4 weeks.

Bacterial liquid cultures were all grown in L-broth unless otherwise stated.

II.2.C. Growth and storage of bacteriophage lambda strains

Bacteriophage lambda stocks were stored at 4°C in 5 ml of phage buffer (10 mM tris-HCl pH 7.5, 100 mM NaCl, 50 mM

Table 2

Strain	Genotype	Source	Reference
i) <i>E. coli</i> strains:			
1803	ra ⁻ , m ⁺ , azi ⁻ , met ⁻	C Wilson, Warwick	Wood, 1966
DM1	F ⁻ , recA1, <u>endA1</u> , <u>avrA96</u> , thi-1, <u>hmd817</u> (ra ⁻ , m ⁺), <u>supE44</u> , <u>relA1</u> , <u>λ</u> ⁻	S Sheikh, Warwick	Nanshan, 1983
JN101	(<u>lac</u> <u>pro</u>), thi, <u>his</u> , <u>supE</u> , <u>endA</u> , <u>hmd</u> , <u>sup</u> , <u>λ</u> <u>tsa36</u> , <u>recA8</u> , <u>azi</u> , 2 Mit	S Sheikh, Warwick	Messing, 1979
DM5<	F-m80d1acZAM13, Δ(lacZYA-araF)U169, <u>recA1</u> , <u>endA1</u> , <u>hmd817</u> (ra ⁻ , m ⁺), <u>supE44</u> , <u>λ</u> ⁻ , thi-1, <u>avrA</u> , <u>relA</u>	R Spooner, Warwick	Focus, Spring 1986, p19
ii) M13 bacteriophages:			
mp18, mp19		R Jenkins, Warwick	Messing, 1984
mp9		Gibco-BRL	Messing and Vieira, 1982
iii) Lambda bacteriophages:			
Charon 35		D Griffiths, Warwick	Loenen and Blattner, 1983
iv) <i>E. coli</i> plasmids:			
pUC 8		S Sheikh, Warwick	Vieira and Messing, 1982
pGEN-blue3, pGEN-blue4		M May, Warwick	Emmenga Biotech Catalogue
pBCL411, pBCL521		LM Roberts, Warwick	Laub et al., 1985, Roberts et al., 1985

Table 2 Bacteria, bacteriophages and plasmids used

Table 1 lists the various E. coli strains, bacteriophages and plasmids used during the course of this study. Where appropriate, the essential genotypic features are detailed. The E. coli plasmids and M13 bacteriophages used were originally obtained as double stranded DNA preparations.

MgCl₂) containing 0.3% (v/v) chloroform (Maniatis et al., 1982).

Where fresh lysates were required, bacteriophage stocks were used to inoculate plating bacteria. Plating bacteria were prepared by growing an overnight culture from a single colony of the appropriate host strain, pelleting the bacterial cells by centrifugation at 2 000 g, 4°C for 10 min and resuspending the cells in a quarter volume of ice cold 10 mM MgCl₂. 0.3 ml aliquots of the plating bacteria in Eppendorf tubes were inoculated with the desired quantity of bacteriophages. The cells were then incubated at 37°C for 15 min and each aliquot added to 3 ml of molten top agarose (L-broth supplemented with 10 mM MgCl₂, 0.4% (w/v) maltose and 0.7% (w/v) agarose). After brief mixing each sample was poured onto a dried L-agar plate (85 mm diameter). Once the agarose had set, the plates were inverted and incubated at 37°C overnight. 5 ml of phage buffer was then added to each plate and the bacteriophages were eluted from the top agarose by agitation on a rotating table for 4 h at room temperature. For optimal bacteriophage yields, a density of 100 000 plaques per plate was aimed for. The liquid containing the eluted bacteriophages was removed from each plate and chloroform was added to 0.3% (w/v). The preparations were stored at 4°C as before. Where required, the bacteriophage concentration of a lysate was determined by

preparing a dilution series in phage buffer and plating out 10 ul of each dilution as before. The number of plaques on each plate could then be counted and the figures used to calculate the bacteriophage concentration in the undiluted lysate.

II.2.D. Growth and storage of M13 bacteriophages

M13 bacteriophages were stored either as single stranded DNA or double stranded RF (replicative form) DNA. In both cases the DNA was stored at 20°C dissolved in sterile distilled water.

Growth of M13 bacteriophages was performed as described in section 12.

SECTION II.3. ISOLATION OF CASTOR BEAN NUCLEAR DNA

The protocol described here was devised with the help of D Griffiths and is based in part upon methods previously described (Gallagher and Ellis 1982, Kislev and Rubenstein 1980)

II.3.A. Isolation of nuclei

Seedlings for the extraction of nuclei were grown under the regime previously described. The cotyledons, still etiolated and enclosed within the seed endosperm, and sometimes the testa were removed by hand. The cotyledons were then kept on ice in a glass beaker until enough tissue had been harvested for homogenisation. Cotyledons were chopped crudely with a razor blade and 2 ml of an extraction buffer EB+ (0.44 M Sucrose, 2.5% (w/v) ficoll 400, 0.4% (w/v) Dextran T40, 25 mM tris-HCl, 10 mM MgCl₂, 0.5% (v/v) Triton X-100, 10 mM 2-mercaptoethanol, 0.04% (w/v) ethidium bromide, the overall pH adjusted to 7.6 with HCl) was added per gramme of tissue. The mixture was then homogenised for approximately 15 s using a polytron. The resultant slurry was filtered through 2 layers of muslin into a clean beaker. The pulp retained by the muslin was removed and re-homogenised with a further 1 ml of EB+ per g of tissue used. The filtrates from these two stages were then pooled and centrifuged at 4 000 g for 5

min at 4°C. The resultant pellets containing the nuclei were resuspended in approximately 2 ml of EB- (EB+ extraction buffer lacking ethidium bromide) per 100 ml of homogenate. 3-4 ml of this suspension was then loaded onto a percoll step gradient in a 15 ml corex tube. The gradients used consisted of 4 x 2.5 ml layers poured successively in the following order:- 85% (w/v) sucrose; 80% (w/v) percoll; 60% (w/v) percoll; 40% (w/v) percoll. The solutions containing percoll were freshly prepared in a buffer of 0.44 M sucrose, 25 mM tris-HCl pH 7.6, 10 mM MgCl₂. Following centrifugation for 30 min at 4 500 g and 4°C, a band of nuclei was removed from the interface between the 85% (w/v) sucrose layer and the 80% (w/v) percoll layer and resuspended in 15 ml of EB- in a fresh 15 ml corex tube. This suspension was centrifuged for 5 min at 4 500 g and 4°C and the pellet resuspended in 15 ml EB-. The centrifugation step was repeated as before and the final pellet resuspended in 1 ml of nuclei storage buffer (0.44 M sucrose, 50 mM tris-HCl pH 7.8, 5 mM MgCl₂). 10 ul of the resulting preparation was then added to 490 ul of nuclei storage buffer and a nuclei count made using a haemocytometer. If the nuclei were not to be used immediately for DNA preparation, they were then transferred to a 1.5 ml Eppendorf tube and stored at -80°C. Preparation of DNA from frozen samples normally took place within a few days.

II.3.B. Isolation of DNA from castor bean nuclei

Preparations of castor bean nuclei were thawed where appropriate and centrifuged for 2 min in a microcentrifuge. After discarding the supernatant the pellet from each 1.5 ml Eppendorf tube was resuspended in a total of 2 ml of DIM A (0.2 M sucrose, 10 mM tris-HCl pH 7.4, 2 mM CaCl_2) and transferred to a 15 ml stoppered polypropylene tube. To the resuspended nuclei was added 3.3 ml of 5 M NaCl, 40 μl of 10 mg/ml ethidium bromide, 1.3 ml of DIM B (10 mM tris-HCl pH 8.5, 10 mM EDTA, 10 mM NaCl, 0.5% (w/v) SDS), 66 μl of 20 mg/ml proteinase K and 134 μl of 2-mercaptoethanol. After mixing, the preparation was incubated for 1 h at 50°C. Another 134 μl of 2-mercaptoethanol was then added prior to a further incubation for 1 h at 50°C. The mixture was then dialysed against 2 sequential 4 l changes of TNE buffer (20 mM tris-HCl pH 8.5, 0.1 M NaCl, 1 mM EDTA), each lasting a minimum of 10 h. Dialysis tubing was prepared by boiling lengths of tubing in 20 mM EDTA for 10 min. After removal from the dialysis bag, the preparation was centrifuged at 4 000 g for 5 min to remove any insoluble polysaccharide. The resultant supernatant was then extracted by gentle rolling and inversion only with an equal volume of 1:1 (v/v) TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA)- equilibrated redistilled phenol: chloroform mixture. Centrifugation at 4 000 g for 5 min was then carried out in order to

separate the aqueous and organic phases. The aqueous phase was removed by means of a wide bore Pasteur pipette into a fresh tube and a further 3 extractions carried out. By this stage a transparent preparation was obtained.

II.3.C. CsCl gradient purification of DNA

The volume of the DNA preparation obtained as above was increased to 28 ml by the addition of an appropriate volume of TE buffer. 28 g of CsCl was then added and allowed to dissolve by gentle mixing. Finally, 3 ml of 10 mg/ml ethidium bromide was added and the preparation again gently mixed. The resulting solution was then loaded into a 37 ml Beckman Quickseal tube. Balancing of pairs of tubes where necessary was carried out by the replacement of small volumes in the tubes with either sterile distilled water or sterile CsCl-saturated water. The tubes were then sealed and centrifuged at 45 000 rpm and 20°C in a Beckman Vti50 rotor for a minimum of 17 h. The rotor was brought to rest without the aid of the brake. The resulting DNA band was visualised under long-range uv and withdrawn after puncturing the tubes at the top, with an 18 gauge needle. The ethidium bromide was removed from the DNA sample by 4 extractions with an equal volume of amyl alcohol. The CsCl was removed from the sample by dialysis against 3 sequential 4 l changes of 2 mM tris-HCl pH 8.0, 0.2 mM EDTA, each lasting at least 6 h. A 5-fold

concentration of the sample was then achieved by wrapping the tubing with aquacide. This caused water to pass across the membrane thus reducing the sample volume. It was necessary to replace the layer of aquacide a number of times in order to achieve the required reduction in volume. Finally the sample was carefully removed from the tubing and stored in an Eppendorf tube at 4°C prior to use. The DNA obtained was found to be of sufficiently high molecular weight (up to approximately 70 Kb) for both Southern blotting and genomic cloning. Both the size and quantity of the DNA obtained were assessed by agarose gel electrophoresis.

SECTION II.4. SOUTHERN BLOTTING AND HYBRIDISATION OF GENOMIC DNA

II.4.A. Restriction endonuclease digestion

Digestion of castor bean nuclear DNA with restriction endonucleases was carried out using the 10 x buffers supplied by Amersham International PLC with its enzymes. In order to obtain enough digested DNA for a single track on the agarose gels shown, 10 ug of nuclear DNA was restricted in a reaction volume of 200 ul. A 5-fold excess of restriction enzyme (normally approximately 50 u) was used and additionally spermidine HCl was added to a final concentration of 4 mM. Digestion was carried out at 37°C for a minimum of 3 h. Following digestion the DNA was precipitated by the addition of 20 ul of 3 M sodium acetate pH 4.8 and 500 ul ethanol. After chilling the samples on dry ice for 30 min, the DNA was pelleted by centrifugation for 10 min in a microcentrifuge. The pellets were then dried in vacuo for 5 min and each resuspended in 20 ul TE.

II.4.B. Agarose gel electrophoresis

For high resolution of genomic DNA fragments, 0.6% (w/v) agarose gels were made and run in 1 x TAE buffer (0.04 M tris-acetate pH 8.2, 1 mM EDTA) as described by Maniatis

et al (1982). The appropriate weight of agarose was dissolved in 1 x TAE using a microwave oven at 10 W power for 7 min. The gel was cooled to 60°C before pouring. Leicester Biocentre gel tanks were used; 10 cm or 15 cm gels were poured using either 200 ml or 350 ml of 1 x TAE agarose. The combs used for such gels contained 12 mm wide slots. The gel was left to set for at least 1 h, then placed in the gel tank and covered in either 1.5 l or 2.5 l of 1 x TAE buffer, depending on the gel size. The samples were mixed with 5 ul 5 x loading buffer (50% (v/v) glycerol, 1% (w/v) ficoll 400, 0.1% (w/v) SDS, 25 mM EDTA) and loaded into the wells. A sample containing Hind III - digested bacteriophage lambda DNA size markers, both ³²P-labelled and unlabelled, was also loaded onto the gel in the same way. 10 cm gels were run at 35 V for 16 h and 15 cm gels at 60 V for 18 h. The gels were stained in 1 ug/ml ethidium bromide for 30 min and were visualised on an ultra violet transilluminator. Gel photographs were obtained using a Polaroid instant camera with Polaroid 665 film.

II.4.C. Transfer of DNA onto hybridisation membranes

This was performed using the following modified procedure of Southern (1975). Following visualisation under uv, gels to be transferred onto hybridisation membrane were first treated with 500 ml of depurinating solution (0.25 M

HCl) in a plastic box, agitated using a rotating table. This process improves the transfer efficiency of larger DNA fragments (Wahl et al, 1979). Depurination was allowed to proceed for 20 min. The liquid was then poured off and the gel treated in the same way with 500 ml of 1.5 M NaCl, 0.5 M NaOH to bring about denaturation of the DNA. After 20 min the denaturing solution was replaced with another 500 ml of fresh solution and the gel treated for a further 20 min. The liquid in the box was then removed and replaced with 500 ml of 1.5 M NaCl, 0.5 M tris-HCl pH 7.4 to bring about neutralisation of the gel. Again this was allowed to proceed for a total of 40 min, the solution being replaced with the same volume of fresh solution after 20 min. The DNA was transferred onto nitrocellulose or nylon membrane by capillary blotting. This was achieved by first placing the gel onto a platform overlain with Whatman 3MM paper dipping into a reservoir of 20 x SSC (3 M NaCl, 0.3 M tri-sodium citrate pH 7.0). Areas on the platform surrounding the gel were covered with cling film to prevent "short circuiting" of the liquid flow. The transfer membrane was then placed on top of the gel; nitrocellulose membranes were pre-moistened with 2 x SSC whereas nylon membranes were not pre-moistened. 2 layers of Whatman 3MM paper cut to the same size as the membrane were placed over its exposed surface, followed by paper tissues to a depth of 4 cm. Finally the tissues were overlain with a glass plate on top of which

was placed a weight of approximately 1 kg. Transfer of the DNA was allowed to proceed for 12-18 hours. Following transfer of the DNA, membranes were briefly washed in 2 x SSC to remove any adhering agarose and allowed to dry for 30 min. The DNA was then fixed to the membrane by either baking in vacuo for 2 h at 80°C in the case of nitrocellulose or exposing the membrane (the side which came into contact with the gel) on an ultra violet transilluminator for 5 min in the case of nylon membranes.

II.4.D. Hybridisation of radiolabelled probe to immobilised DNA

Two different hybridisation protocols were used, depending on the type of transfer membrane employed:

II.4.D.1) Nitrocellulose membranes

A method modified from that of Maniatis et al (1982) was employed. Membranes were first pre-hybridised in a sealed bag for 2-3 h at 42°C in 100 ul of prehybridisation buffer per cm² of membrane. The pre-hybridisation buffer consisted of:- 1 x SSC, 5 x Denhardt's solution (100 x Denhardt's solution is 2% (w/v) polyvinylpyrrolidone-40, 2% (w/v) ficoll 400, 2% (w/v) bovine serum albumin), 50% (v/v) formamide (deionised by stirring with amberlite MB-3 monobed resin), 0.1% (w/v) SDS and 100 ug/ml polyadenylic acid). Following prehybridisation, the radiolabelled probe, previously denatured in a boiling water bath for 5

min and transferred directly to ice water for 1 min, was added to the bag. The bag was resealed and the probe thoroughly mixed into the prehybridisation solution. Radiolabelled probes with specific activities of typically $1-5 \times 10^7$ cpm per μg of DNA were used for hybridisation to Southern blots of this type. Hybridisation was carried at 42°C for 16-20 h.

11) Nylon membranes

The procedure used was modified from that recommended by Amersham International PLC for use with Hybond-N membranes. Prehybridisation was carried out in a sealed bag for 2-3 h at 65°C in 50 μl of prehybridisation buffer per cm^2 of membrane. The prehybridisation buffer used consisted of 1- 6 x SSC, 5 x Denhardt's solution, 0.5% (w/v) SDS and 100 $\mu\text{g}/\text{ml}$ polyadenylic acid. The radiolabelled probe was denatured and added to the prehybridisation solution as described in the method used for nitrocellulose membranes. Hybridisation was carried out at 65°C in an otherwise similar manner.

Following hybridisation, membranes were washed 3 times at room temperature in 2 x SSC, 0.1% (w/v) SDS for 10 minutes. They were then washed further for 60 min either in 1 x SSC at 55°C (low stringency washes) or in 0.1 x SSC at 65°C (high stringency washes). Subsequently, the membranes were sealed in bags and exposed for 1-7 days to

Kodak X-omat S X-ray film in conjunction with a Cronex intensifying screen at -80°C .

SECTION II.5. CONSTRUCTION OF GENOMIC LIBRARIES IN
BACTERIOPHAGE LAMBDA

Two genomic libraries were constructed in the bacteriophage lambda vector charon 35 (Loenen and Blattner, 1983), one from partial Eco RI digests of castor bean DNA, the other from partial Sau 3A digests of castor bean DNA. Both libraries were constructed using the method described by Maniatis et al (1982). For convenience, the description below will refer only to the partial Sau 3A library, since the same procedure was followed for the construction of the partial Eco RI library, the only difference being the restriction enzymes used. In the latter case both vector and genomic DNAs were cleaved using Eco RI.

II.5.A. Determination of optimal conditions for the
production of 10-20 Kb genomic fragments

Initially, 24 ug of purified castor bean DNA was gently mixed with 72 ul of 10 x Sau 3A buffer and sufficient sterile distilled water to create a total volume of 720 ul. 120 ul of this mixture was then placed in an Eppendorf tube (labelled "tube 1") and 60 ul aliquots were removed separately into 10 other Eppendorf tubes (labelled 2-11). All tubes were chilled on ice for 10 min. 3.2 ul of 10 u/ul Sau 3A was then added to tube 1. Following

gentle mixing a 60 ul aliquot was removed from this tube and placed in tube 2. Gentle mixing was again carried out and a 60 ul aliquot removed from tube 2 into tube 3. This process was repeated along the line of tubes until finally the 60 ul aliquot removed from tube 10 was discarded. No enzyme was added to tube 11, but it was mixed to the same degree as the other tubes and served as a control by which to assess the extent of mechanical shearing of the DNA. Following mixing the tubes were incubated at 37°C for 1 h. They were then immediately transferred onto ice and mixed with 2.4 ul of 0.5 M EDTA in order to prevent any further restriction endonuclease activity. 15 ul of 5 x loading buffer was added to each sample, as well as a sample of uncut castor bean DNA not subjected to the mixing process. All samples were electrophoresed overnight on a 0.5% (w/v) agarose, 1 x TAE gel (with specially enlarged loading slots) in the presence of digested lambda DNA size markers. Following ethidium bromide staining, the gel was visualised and the DNA sample with the greatest fluorescence peak in the size range 10-20 Kb noted.

II.5.B. Large scale preparation of genomic fragments for cloning

The optimal conditions previously noted were subsequently used in the large scale preparation of genomic fragments for cloning. Furthermore, to increase the chances of

cloning any given sequence of interest, large scale digests were also carried out using the enzyme:DNA ratios immediately above and below the optimum. Since restriction sites are non-randomly distributed in any given genome and frequently show non-uniform susceptibility to the enzyme in question, such a precaution is likely to enable the production of a more random population of cloned DNA fragments. Thus large scale digests were carried out using the enzyme: DNA ratios of 0.12 u/ug, 0.06 u/ug and 0.03 u/ug. All conditions used were as before except that the reaction volume was increased to 600 ul. The size distribution of the digested DNA was checked by gel electrophoresis as before. The DNA was then gently extracted twice with 1:1 (v/v) TE-equilibrated phenol/chloroform mixture, ethanol precipitated as per normal and pelleted by centrifugation at 10 000 g for 15 min. The precipitate was washed once with 70% (v/v) aqueous ethanol, dried in vacuo and resuspended in 500 ul of TE.

II.5.C. Fractionation of genomic fragments

For optimal ligation conditions it was next necessary to isolate clonable genomic fragments (ie 10-22 Kb) from the total pool of fragments generated in the partial digests. This was achieved by sucrose density gradient centrifugation. The DNA sample (250 mg in total) was

heated for 10 min at 70°C, cooled to 20°C and loaded onto a 38 ml, 10-40% (w/v) sucrose density gradient in a Beckman SW 28 Ultraclear tube. The sucrose solutions used contained, in addition to the sucrose, 1 M NaCl, 20 mM tris-HCl pH 8.0 and 5 mM EDTA. The gradient was centrifuged at 26 000 rpm for 24 hours at 20°C in a Beckman SW28 rotor and brought to rest without use of the brake. Fractions were removed by puncturing the bottom of the tube and collecting 0.5 ml samples. To 30 ul of every third fraction was added 30 ul of sterile distilled water and 15 ul of 5 x loading buffer. These samples were then analysed by gel electrophoresis on a 0.4% (w/v) agarose, 1 x TAE gel (with enlarged sample slots) using digested bacteriophage lambda DNA size markers. The sucrose and NaCl concentrations of the size marker samples were adjusted to correspond with genomic DNA samples in the middle of the gel. Fractions containing DNA within the desired size range were pooled and dialysed against 3 sequential 4 l changes of TE buffer, each lasting a minimum of 12 h. The DNA was then extracted with butan-1-ol in order to reduce the volume to approximately 1 ml. The DNA was ethanol precipitated and pelleted using a microcentrifuge. After drying in vacuo, the DNA was redissolved in TE at a concentration of approximately 0.3 ug/ul. The purified genomic fragments were 5'-dephosphorylated prior to ligation with the lambda arms so that the possibility of double insertion events would be

minimised. The procedure described in section II.8.A. was used for this purpose. After 5'-dephosphorylation, the fragments were again ethanol precipitated and redissolved in TE at a concentration of approximately 0.3 ug/ul. The exact concentration and size distribution of the purified genomic fragments were established by electrophoresis using high and low percentage agarose gels respectively.

II.5.D. Preparation of bacteriophage lambda arms

In order to generate bacteriophage lambda arms with compatible ends for cloning Sau 3A-digested DNA, the lambda vector was cleaved with Bam HI restriction enzyme. 50 ug of DNA prepared from the bacteriophage lambda strain charon 35 was digested in a reaction volume of 500 ul with 150 u of Bam HI in the usual way. An aliquot of 5 ul was removed and heated at 70°C for 10 min to disrupt the cohesive ends of the DNA. This aliquot was then analysed by gel electrophoresis through a 0.5% (w/v) agarose, 1 x TAE gel to check that the DNA had been completely digested. The sample was extracted twice with an equal volume of TE-equilibrated 1:1 (v/v) phenol: chloroform mixture and the DNA precipitated by the addition of 50 ul 3 M sodium acetate pH 4.8 and 500 ul ethanol. The DNA was pelleted in a microcentrifuge, dried in vacuo and resuspended in 333 ul of TE. 3.6 ul of 1 M $MgCl_2$ was then added and following mixing, the tube was incubated at 42°C

for 1 h to allow the bacteriophage lambda cohesive ends to anneal. Again a 5 ul aliquot was removed for analysis by gel electrophoresis as before. A further 5 ul aliquot, denatured as previously described, was used as a size marker on the same gel. Reannealing of the cohesive ends was observed by the appearance of a 30 Kb DNA fragment. The annealed arms were purified on a 10-40% (w/v) sucrose density gradient using the same procedure as was described for the purification of 10-20 Kb genomic fragments. In this case a 15 ul aliquot was removed from every third fraction and diluted with 35 ul of sterile distilled water. 8 ul of loading dye was then added and the samples heated at 70°C for 10 min. Fractions containing the annealed arms were identified by electrophoresis through a 1 x TAE, 0.5% (w/v) agarose gel. These showed a visible separation from fractions containing either the unannealed left arm or right arm, which themselves occurred in distinct fractions. Since the stuffer region of the charon 35 vector is cleaved by Bam HI at several restriction sites, there are a number of smaller stuffer fragments generated, all of which showed good separation from the arms. Fractions containing the annealed arms were pooled, dialysed, butan-1-ol extracted, ethanol precipitated, washed and redissolved in TE as previously described for the purified genomic fragments. The yield and purity of the preparation was assessed by electrophoresis through a 0.5% (w/v) agarose, 1 x TAE gel.

II.5.E. Ligation of genomic fragments to lambda arms

By considering the average molecular weight of the purified genomic fragments and the respective molecular weights of the two bacteriophage lambda arms, it would be expected that the optimum ratio (w/w) of arms:inserts for ligation would be 2:1. However, some molecules may lack a cohesive terminus, thereby altering the effective concentration of ends available for ligation. Thus 3 different ligation reactions were assembled using the ratios 1:1, 2:1 and 3:1 (w/w) of arms:inserts. In this way it was hoped that the optimum ligation conditions would be achieved in at least one of the reactions. The relative efficiencies of the ligation reactions could later be observed by their corresponding yields of bacteriophages. Each ligation was carried out in a reaction volume of 10 ul containing a total of 2 ug of DNA. The high concentration of DNA is desirable since it maximises the formation of bacteriophage lambda DNA concatemers. Such a configuration is the optimum for in vitro packaging. The DNAs to be ligated were assembled in a small Eppendorf tube and 1 ul of freshly-prepared 10 x ligation buffer (0.66 M tris-HCl pH7.5, 50 mM MgCl₂, 50 mM dithiothreitol, 10 mM ATP) added. Sterile distilled water was then added to make a total volume of 10 ul. Finally, 0.5 ul of 2.5 u/ul T4 DNA ligase was added, the components mixed and the tubes incubated for at least 16 h at 15°C.

II.5.F. Packaging of DNA and titration

This was performed according to the manufacturer's recommendations. The packaging extracts, previously stored at -80°C , were allowed to thaw on ice. To each extract was added the contents of a single ligation reaction. Gentle mixing was achieved by means of a sealed capillary tube. The packaging reactions were carried out by incubating the tubes at 22°C for 2 h. 250 μl of 20 mM tris-HCl pH 7.9, 20 mM NaCl, 10 mM MgSO_4 was then added, followed by 20 μl of chloroform. After gentle vortexing of the tubes, the yield of bacteriophages was assayed by preparing 10^{-1} , 10^{-2} and 10^{-3} dilutions of the final product and plating out 10 μl of each of these. The E. coli strain K803 was used for this purpose and the preparation, adsorption and plating procedures were carried out as previously described. The size of genomic library obtained could thus be determined. The percentage of non-recombinants in the genomic libraries constructed for this study was assumed to be in the region of 5-10%, based on calculations from the data of Murray et al. (1984).

SECTION II.6. SCREENING OF GENOMIC LIBRARIES FOR LECTIN SEQUENCES

II.6.A. Growth of bacteriophage lambda plates for screening

In order to maximise the chances of detecting any given cloned gene of interest, bacteriophages derived from in vitro packaging were plated out and the resultant plaques screened directly. Plaques were grown on square 10cm x 10 cm Sterilin plates at a density of 250 pfu/cm² using the E. coli strain K803 as the host. A total of 9×10^5 plaques from the partial Sau 3A library were screened, as well as the entire partial Eco RI library. The direct screening approach used ensured that problems associated with unequal growth rates of recombinant clones were kept to a minimum. Plates were grown overnight at 37°C and chilled at 4°C for a minimum of 1 h prior to plaque lifting.

II.6.B. Plaque Lifts

Bacteriophage DNA was transferred onto nitrocellulose filters in the form of a plate replica using the standard technique of Benton and Davis (1977), as described by Maniatis et al (1982). Sheets of nitrocellulose cut to the size and shape of the agar plates were labelled on one

side using a red ballpoint pen and a distinguishing mark was made along one of the four edges. A similar mark was made on one edge of each plate to be screened. Each nitrocellulose filter was placed carefully on top of its corresponding plate with the distinguishing marks aligned together. After 1 min the filter was removed by means of forceps and immersed, DNA side up, in a shallow tray of denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 1 min. The filter was then placed in a shallow tray of neutralising solution (1.5 M NaCl, 0.5 M tris-HCl pH 8.0) for 5 min. Finally, each filter was rinsed in 2 x SSC and placed on Whatman 3MM paper to dry. The DNA was fixed to the nitrocellulose by baking in vacuo for 2 h at 80°C.

II.6.C. Hybridisation of bacteriophage plaques to radiolabelled cDNA probe

Again the procedure used was based on the method of Benton and Davis (1977), as described by Maniatis et al (1982). Prehybridisation and hybridisation of the filters was carried out in an identical fashion to that described for Southern blots on nitrocellulose (section 2.4.D.). The ³²P-radiolabelled probe used was obtained by nick translation of the ricin cDNA insert in the plasmid pRCL611 (Lamb, 1984). The DNA fragment in question was excised from the recombinant plasmid using the enzyme Pst I. The cDNA insert lacks at its 5' end the 5'

untranslated region and the first 3 bp encoding the ricin signal sequence. Also lacking at the 3' end are the terminal 252 bp of the ricin B-chain coding sequence and 3' untranslated region. Following hybridisation, the filters were washed 3 times at room temperature in 2 x SSC, 0.1% (w/v) SDS for 10 minutes. They were then washed further for 60 min in 1 x SSC at 55°C, sealed in cling film and mounted with their distinguishing marks uppermost onto sheets of Whatman 3MM paper. A number of labels distinctively marked with radioactive ink were placed around the filters to allow subsequent alignment between the filters and autoradiographs.

Autoradiography was carried out overnight at -80°C using Kodak X-omat S X-ray film in the presence of 2 Cronex intensifying screens. Lectin-positive clones were identified as defined black spots on the autoradiographs. By means of the alignment marks previously described, it was then possible to remove a core of top agarose, from the plate in question, corresponding to the position of the positive signal. The bulb end of a wide bore Pasteur pipette was used for this purpose. To the core of top agarose removed was added 1 ml of phage buffer and 50 ul of chloroform in an Eppendorf tube. After vortexing, the bacteriophages were left to elute overnight at 4°C. Again the tubes were vortexed and a dilution series (from 10^{-2} to 10^{-4}) was made from each bacteriophage eluate. 10 ul

of each dilution was plated out and the resultant plaques counted. These figures were then used to calculate the bacteriophage concentration of the original eluate. This titration was carried out using the E. coli strain DH1, which was used for all subsequent screening and titration of bacteriophages. The same procedure was followed for each round of screening using a progressively lower plaque density until it became possible to identify individual positive plaques on the plates themselves. This was achieved with most clones after 3 rounds of screening. After individual positive plaques had been picked using a Pasteur pipette, one further round of screening was carried out to ensure that plaque purity had been attained.

SECTION II.7. PREPARATION AND ANALYSIS OF DNA FROM
POSITIVE CLONES

II.7.A. Growth of bacteriophages and extraction of DNA

Three methods were used to isolate bacteriophage lambda DNA. The first method described below yielded DNA of sufficient quality for Southern blotting analysis only, whereas the second and third methods provided DNA which was also suitable for subcloning.

II.7.A.1) Rapid, small scale preparation of bacteriophage
lambda DNA

It was found that the following unpublished method (C. Bowler, personal communication) could be used to prepare DNA from all 17 bacteriophage lambda clones simultaneously. An individual plaque of each clone was picked from a freshly grown plate using a Pasteur pipette and the bacteriophages eluted in 500 ul of 10 mM tris-HCl pH 7.2, 10 mM MgSO₄, 0.02% gelatin for 30-60 min at room temperature. 50 ul of the eluate was then mixed with 50 ul of 10 mM MgCl₂, 10 mM CaCl₂ and 50 ul of an E. coli strain K803 culture in late-logarithmic growth phase. This mixture was incubated for 10 minutes at room temperature to allow the bacteriophages to adsorb to the bacterial cells and was then added to 4 ml of MZ medium (1% (w/v) NZamine, 85 mM NaCl, 10 mM MgCl₂) in a 25 ml

conical flask. The NZamine (a casein hydrolysate) was kindly provided by C Bowler. The flasks were shaken vigorously at 37°C and 300 rpm for approximately 6 h or until complete cell lysis was apparent. 2 ml of each lysate was then transferred to a stoppered 10 ml polypropylene tube and 0.4 ml of 2.5% (w/v) SDS, 0.25 M EDTA, 0.5 M tris-HCl pH 9.0 was added. After vortexing, the tubes were incubated at 70°C for 30 min. 0.8 ml of 5 M potassium acetate was then added and the tubes placed on ice for 15 min, followed by centrifugation at 9 000 g for 20 min. The resultant supernatants were poured off through one layer of Whatman number 1 paper into fresh tubes, their volumes measured and the nucleic acids precipitated by the addition of 2 volumes of ethanol. After chilling at -80°C for 20 min, the tubes were centrifuged at 9 000 g for 30 min. The resultant pellets were resuspended in 0.4 ml of 2 M ammonium acetate and the preparations transferred to Eppendorf tubes. Any remaining insoluble material was removed by pelleting in a microcentrifuge for 5 min, after which the supernatants were decanted into fresh tubes and their volumes measured. 2 volumes of ethanol were added to each tube and the preparations chilled at -70°C for 20 min. After the nucleic acids had been pelleted by centrifugation for 10 min in a microcentrifuge and dried in vacuo, 50 ul of sterile distilled water was added and the pellets redissolved. The preparations were stored at -20°C prior

to use. Each normally contained sufficient DNA for 5 restriction digests.

II.7.A.ii) Bacteriophage lambda DNA "midipreparations"

In order to prepare DNA from the bacteriophage lambda clones of sufficient quality for subcloning, the method of Kao et al (1982) was used. 50 ml quantities of prewarmed LB medium supplemented with 5 mM MgSO_4 were each inoculated with 0.5 ml of an overnight K803 culture in a 500 ml conical flask. The flasks were then shaken at 37°C, 200 rpm for approximately 2 hours until an A_{600} of 0.3 was reached. Each flask was inoculated with 10 bacteriophages from a freshly prepared and titred lysate and shaken at 37°C, 300 rpm until complete lysis of the bacterial cultures was apparent. This usually took at least 4 hours. 0.5 ml of chloroform was added to each lysed culture and the flasks were shaken for a further 5 min at 200 rpm, room temperature. The cellular debris were removed by centrifugation for 15 minutes at 4 000 g and 4°C. To each supernatant was added 50 ul of 10 mg/ml RNase A and 50 ul of 10 mg/ml DNase I. Digestion was allowed to proceed overnight at 4°C. The bacteriophages were then pelleted by centrifugation for 1 h at 20 000 g and 4°C, and each pellet resuspended in 0.3 ml of 0.1 M tris-HCl pH 7.9, 0.3 M NaCl, 10 mM MgCl_2 . The preparations were left overnight at 4°C and each transferred to an Eppendorf tube. Following 3 extractions

with an equal volume of TE-equilibrated 1:1 (v/v) phenol/chloroform mixture, the bacteriophage DNA was precipitated with 2 volumes of ethanol. After the preparations had been chilled for 30 min on dry ice, the tubes were centrifuged for 10 min in a microfuge to pellet the DNA. Pellets were then washed once with 70% (v/v) aqueous ethanol, dried in vacuo and each redissolved in 50 μ l of sterile distilled water. Using this method yields of approximately 5 μ g per preparation were normally obtained.

II.7.A.iii) Large scale preparation of bacteriophage
lambda DNA

The method of Maniatis et al (1982) was followed where the preparation of larger quantities of high quality DNA was necessary. This method is not well suited for use in isolating DNA simultaneously from several bacteriophage lambda clones and therefore was used to prepare DNA of the charon 35 vector and lambda clone 1 only. Initially a single colony of E. coli strain DH1 was inoculated into 10 ml of LB. The culture was shaken at 37°C, 200 rpm overnight. Aliquots of 1 ml were then removed and each added to 250 ml of SB (3.5% (w/v) bacto-tryptone, 2% (w/v) yeast extract, 85 mM NaCl, 10 mM MgSO₄) in a 2 l flask. Normally 6 cultures were inoculated in this way for each bacteriophage lambda clone to be grown. The cultures were shaken at 37°C, 250 rpm until an A₆₀₀ value of 0.3 was

reached. Each flask was then inoculated with 1 ml of phage buffer containing bacteriophages of the desired strain. Normally a different quantity of bacteriophages, ranging from 5×10^6 pfu to 10^9 pfu, was added to each culture. The flasks were shaken at 37°C , 300 rpm overnight. After at least 16 h, 5 ml of chloroform was added and the cultures shaken as above for a further 30 min. Bacterial nucleic acids were removed by the addition of 25 μl of 10 mg/ml RNase A and 25 μl of 10 mg/ml DNase I to each flask, followed by incubation at room temperature for 30 min. To each flask was then added 14.6 g of NaCl, which was dissolved by intermittent agitation at room temperature for 1 h. Each 250 ml sample was then poured into an MSE 300 ml centrifuge bottle and centrifuged at 11 000 g, 4°C for 10 min. The supernatants were each poured into a 500 ml flask to which 25 g of solid PEG 6000 was added. The preparations were mixed using magnetic stirring bars and then placed on ice for 1 h once the PEG 6000 had completely dissolved. The resultant bacteriophage precipitates were collected by centrifugation as above. The pellets were drained well and each resuspended in 2 ml of phage buffer. The preparations were then pooled in an MSE 40 ml polypropylene Oakridge tube and extracted once against an equal volume of chloroform. CsCl was added to the bacteriophage suspension to a final concentration of 0.5 g/ml. The preparation was then layered onto CsCl step

gradients prepared in 14 ml Beckman Ultra-clear centrifuge tubes. The gradient solutions were made in phage buffer and consisted of 1 ml of 1.27 g/ml CsCl, 1.5 ml of 0.817 g/ml CsCl and 1.5 ml of 0.706 g/ml CsCl. The gradients were centrifuged in a Beckman SW40Ti rotor at 22 000 rpm and 4°C for 2 h. The bacteriophage bands, which form at the interface between the two uppermost CsCl solutions were removed using a 5 ml syringe with an 18 gauge needle. The samples were pooled in a 5 ml Beckman Ultra-clear centrifuge tube which was then topped up with 1.5 g/ml CsCl in phage buffer. The preparation was centrifuged in a Beckman SW50.1 rotor for 24 h at 35 000 rpm and 4°C in order to establish an equilibrium density gradient within the tube. The resultant bacteriophage band was removed as above and dialysed for several hours at 4°C against 4 l of 10 mM tris-HCl pH 8.0, 10 mM MgCl₂. The dialysis buffer was then replaced with 4 l of fresh solution and further dialysis allowed to proceed overnight at 4°C. DNA was purified from the bacteriophage particles by extraction once with TE-equilibrated phenol and twice with TE-equilibrated 1:1 (v/v) phenol/chloroform mixture. The DNA was precipitated by the addition of 3 M sodium acetate pH 4.8 to 0.3 M and two volumes of ethanol. The precipitate was then collected by centrifugation in a microcentrifuge for 10 min, washed twice with 70% (v/v) aqueous ethanol, dried in vacuo and redissolved in 500 ul of TE. The DNA preparations were stored at 4°C. Yields of DNA obtained

by this method were very variable, but normally in the region of 100-300 ug.

II.7.B. Restriction and Southern blotting analysis of DNA

Lectin-positive genomic clones were initially classified into preliminary groups on the basis of their pattern of hybridisation on Southern blots to a radiolabelled ricin cDNA probe. Two 10 ul aliquots were removed from each DNA minipreparation and separately digested with the restriction enzymes Hind III and Eco RI respectively. 1 ul of 10 mg/ml RNase A, previously boiled for 15 min, was added to each digest. The samples were then electrophoresed on a 0.7% (w/v) agarose, 1 x TAE gel in the presence of digested bacteriophage lambda DNA size markers. The gel was Southern blotted onto nitrocellulose in the manner previously described, and the filter hybridised with the same ricin cDNA probe, again radiolabelled by nick translation, as was used for screening the genomic libraries. Following hybridisation the Southern blot filters were washed 3 times at room temperature in 2 x SSC, 0.1% SDS for 10 minutes. They were then washed further for 60 min in 1 x SSC at 55°C. Autoradiography was carried out as described in section II.4.D.

SECTION II.8. SUBCLONING OF LECTIN GENES AND THEIR
FLANKING SEQUENCES INTO PLASMID VECTORS

II.8.A. Restriction and dephosphorylation of plasmid
vector

The plasmid vector pUC8 (see Appendix I) was used in all cases for the subcloning of DNA fragments excised from the lectin- positive bacteriophage lambda clones. Normally 1 ug of pUC8 DNA was linearised with the appropriate restriction enzyme in the usual way. The digest was then extracted once with an equal volume of TE-equilibrated 1:1 (v/v) phenol/chloroform mixture. The DNA was precipitated by the addition of a tenth volume of 3 M sodium acetate pH 4.8 and 2 volumes of ethanol. After chilling on dry ice for 20 min, the DNA was pelleted by centrifugation for 10 min in a microcentrifuge, dried in vacuo and redissolved in 20 ul of sterile distilled water. Removal of the terminal 5' phosphate groups of the linearised vector was achieved using the method described by Maniatis et al (1982). To the resuspended DNA was added 2.5 ul of 10 x phosphatase salts (10 mM $MgCl_2$, 1 mM $ZnCl_2$), 2.5 ul of 0.5 M glycine (made to pH 9.4 with NaOH) and 1 ul of 1 u/ul calf intestinal phosphatase. The mixture was incubated for 15 min at 37°C, and then for 15 min at 56°C. Following the addition of a further 1 ul of 1 u/ul calf intestinal phosphatase, the incubation procedure was

repeated. 20 ul of sterile distilled water was then added to the preparation, followed by 5 ul of 10 x STE (0.1 M tris-HCl pH 8.0, 1 M NaCl, 10 mM EDTA) and 5 ul of 0.1 M nitrilotriacetic acid. The tube was heated at 68°C for 15 min and 6 ul of 3 M sodium acetate pH 4.8 was added. The preparation was extracted 3 times with an equal volume of 1:1 (v/v) TE-equilibrated phenol/chloroform mixture and 120 ul of ethanol added to precipitate the DNA. The precipitate was then pelleted in a microcentrifuge, washed 3 times with 70% (v/v) aqueous ethanol, dried in vacuo and redissolved in 10 ul of TE. 1 ul of this preparation was electrophoresed on an agarose gel as described in section II.10.B. in order to verify that complete linearisation had taken place, and to estimate the concentration of DNA in the preparation. Dephosphorylated vector DNA was stored at -20°C.

II.8.B. Gel isolation of DNA fragments

This was carried out using a similar procedure to that described by Maniatis et al (1982). Digested DNA was electrophoresed, in this case on a 0.7% (w/v) low melting point agarose, 1 x TBE (89 mM tris, 89 mM boric acid, 20 mM EDTA) gel. Both the gel and the running buffer (also 1 x TBE) used contained 0.5 ug/ml ethidium bromide, thus eliminating the need for staining after electrophoresis. The digested DNA was visualised on a uv transilluminator

and the desired DNA bands identified by means of size markers also present on the gel. A glass plate was placed under the gel and the relevant DNA bands were excised in a minimum volume of agarose using a razor blade. Each slice of agarose was then placed in an Eppendorf tube and 150 μ l of TE added. The agarose was then melted by incubating each tube at 50°C for 10 min and 150 μ l of TE-equilibrated phenol, pre-incubated at 50°C, was immediately added to each tube. The tubes were then centrifuged for 2 min in a microcentrifuge to separate the aqueous and organic phases. A further extraction was carried out using 150 μ l of TE-equilibrated 1:1 (v/v) phenol:chloroform mixture. 3 μ l of 1 μ g/ μ l E. coli tRNA was added to the final aqueous phase and the nucleic acids precipitated by the addition of a tenth volume of 3 M sodium acetate and 2 volumes of ethanol. Finally, after the tubes had been left on dry ice for 20 min, the precipitate was pelleted by centrifugation for 10 min in a microcentrifuge, washed once with 70% (v/v) aqueous ethanol, dried in vacuo and redissolved in 10 μ l of TE. The yield of DNA was assessed in the same way as was described for the dephosphorylated vector. Purified DNA fragment preparations were stored at -20°C.

II.8.C. Ligation of DNAs

DNA fragments isolated from bacteriophage lambda clones were ligated into dephosphorylated linearised DNA of the plasmid pUC8 in a 10 ul reaction mixture containing 1 ul of the same 10 x buffer as was described previously (section II.5.E.). A total of 100 ng of DNA, consisting of vector and insert DNAs in equimolar amounts, was used in each reaction. Incubation was carried out overnight at 15°C.

II.8.D. Transformation of ligated plasmids into E. coli

A method based on that of Mandel and Higa (1970) was used. 50 ml of L-broth in a conical flask was inoculated with 1 ml of an overnight culture of E.coli strain DH5 α and shaken at 37°C until an A₆₀₀ of 0.48 had been reached. The culture was then placed on ice for 5 min prior to centrifugation at 2 000 g, 4°C for 5 min. The resultant cell pellets were gently resuspended in a total of 25 ml of ice-cold 10 mM CaCl₂. The tubes were left on ice for 15 min and then centrifuged again at 2 000 g, 4°C for 5 min. Finally the cell pellets were gently resuspended in a total of 2 ml of ice-cold 50 mM CaCl₂ and kept on ice for a minimum of 30 min prior to use. Aliquots of 200 ul were used for each transformation. Normally 2 transformations were carried for each ligation reaction; one using 5 ul of the ligation mixture and the other using 0.5 ul of the mixture. Transformations were carried out

in separate Eppendorf tubes which were left on ice for 30 min after the competent cells had been inoculated with DNA. The cells were then heat shocked at 42°C for 2 min and 1 ml of L-broth was added to each tube. Following incubation for 1 h at 37°C, the cells were pelleted by centrifugation for a few seconds in a microcentrifuge. Each pellet was then gently resuspended in 100 μ l of L-broth and the cell suspension plated out onto 1.5% (w/v) L-agar plates containing 50 μ g/ml ampicillin. In some cases the agar was supplemented with 0.2 mg/ml IPTG and 0.17 mg/ml X-gal. Under such conditions non-recombinant transformants give rise to blue colonies, whereas colonies from recombinant transformants are white. Thus any non-recombinant clones could be discarded at this stage.

SECTION II.9. PREPARATION OF PLASMID DNA

II.9.A. Rapid, small scale preparations

DNA minipreparations for preliminary restriction endonuclease analysis were isolated using a modification of the alkaline lysis approach of Birnboim and Doly (1979). 5 ml cultures inoculated from the desired colonies were grown in an orbital shaker at 37°C and 200 rpm overnight using L-broth supplemented with 50 ug/ml ampicillin. A 1.5 ml aliquot was removed from each culture into a labelled Eppendorf tube and the cells were pelleted by centrifugation in a microcentrifuge for 1 min. Each pellet was then resuspended in 80 ul of GET (50 mM glucose, 25 mM tris-HCl pH 8.0, 10 mM EDTA). To the suspension was added 5 ul of 40 mg/ml lysozyme (dissolved in GET) and the preparations were left on ice for 5 min. 200 ul of 0.2 M NaOH, 1% (w/v) SDS was then added to each tube and, after thorough vortexing, the preparations were left on ice for a further 10 min. Following the addition of 150 ul of 3 M sodium acetate pH 4.8, again with thorough mixing, the tubes were left for a further 15 min on ice. Chromosomal DNA and assorted debris were then pelleted by centrifugation for 10 min in a microcentrifuge. 0.54 volumes of propan-2-ol was added to each resultant supernatant and the preparations left for 15 min at room temperature, prior to centrifugation for

10 min. The pelleted precipitates were each redissolved as completely as possible in 0.5 ml of 2 M ammonium acetate. A 10 min incubation at 37°C and thorough vortexing were used to achieve this. Any remaining insoluble material was removed by a 10 min centrifugation and 0.54 vols of propan-2-ol was added to each resultant supernatant. The preparations were left for 15 min at room temperature and then centrifuged for 10 min in a microcentrifuge. Finally, the pelleted precipitates were washed once with 70% (v/v) aqueous ethanol, dried in vacuo and each redissolved in 50 ul of TE. 5 ul of this volume was subsequently used per restriction digest. 1% (w/v) agarose, 1 x TBE gels, prepared as previously described, were used for the analysis of digested plasmid DNA by gel electrophoresis. Large scale plasmid DNA preparations were made from clones containing the desired DNA inserts.

II.9.B. Large scale preparations

Again the alkaline lysis method of Birnboim and Doly (1979) was used. 10 ml cultures inoculated from the desired colonies were grown in an orbital shaker at 37°C, 200 rpm overnight using L-broth supplemented with 50 ug/ml ampicillin. 2 l flasks, containing 600 ml of L-broth supplemented with 50 ug/ml ampicillin were each inoculated with a 10 ml overnight culture and grown at 37°C, 200 rpm, overnight. Cells were collected by centrifugation at

7 000 g and 4°C for 5 min in MSE 300 ml polypropylene centrifuge bottles. They were then resuspended in 3 ml of GET per 600 ml culture, transferred to 50 ml MSE Oakridge centrifuge tubes and placed on ice. Cell lysis was achieved by the addition of 200 μ l of GET containing lysozyme at 40 mg/ml and incubation on ice for 5 min. 2 vols of 0.2 M NaOH, 1% (v/v) SDS was added to each tube. After vortexing, the tubes were left on ice for 5 min. A half volume of 3 M sodium acetate pH 4.8 was then added and the contents of the tube briefly vortexed. After a further 30 min period on ice, the tubes were centrifuged at 11 000 g and 4°C for 15 min. The nucleic acids were precipitated by the addition of 0.54 volumes of propan-2-ol to each supernatant in fresh 50 ml Oakridge tubes. Following mixing, the tubes were left at room temperature for 10 min. The precipitates were pelleted by centrifugation at 11 000 g and 4°C for 10 min. The resultant pellets were drained and each redissolved as completely as possible in 3.2 ml of 2 M ammonium acetate by incubation at 37°C for 15 min and vortexing. The samples were then re-centrifuged at 11 000 g, 4°C for 10 min. After centrifugation, 2.1 ml of propan-2-ol was added to each sample in fresh 50 ml Oakridge tubes. The preparations were left for 10 min at room temperature and the resultant precipitates collected by centrifugation as above. Finally the plasmid pellets were washed twice with

5 ml of 70% (v/v) aqueous ethanol, dried in vacuo and resuspended in 1 ml of TE.

II.9.C. Purification of plasmid DNA on CsCl gradients

All procedures up to and including the removal of the DNA band from the CsCl gradient were performed as described for the purification of castor bean nuclear DNA (section II.3.C). The DNA preparations were then transferred to 50 ml polypropylene MSE Oakridge tubes and each volume adjusted to 3 ml by the addition of sterile distilled water. 1.2 ml of 1% (w/v) N-lauroylsarcosine was added to each tube, followed by 1.2 ml of 3 M sodium acetate pH 6.0, 6.6 ml of 10 mM tris-HCl pH 8.0, 0.1 mM EDTA and 24 ml of ethanol. The preparations were chilled overnight at -20°C and the precipitated DNA recovered by centrifugation of the tubes at 11 000 g, 4°C for 15 min. The DNA pellets were then washed 5 times with 70% (v/v) aqueous ethanol, dried in vacuo and redissolved in 1 ml TE. The DNA yield from each preparation was determined by measurement of the A₂₆₀ of a 2 µl aliquot (as per Maniatis et al., 1982).

SECTION II.10. RESTRICTION MAPPING OF PLASMID SUBCLONES

Initially, the subcloned plasmid DNAs were tested for the presence of a number of restriction endonuclease sites by incubation separately with each of the respective enzymes under the appropriate conditions. Having ascertained which restriction sites were present on a given plasmid and how many, the location of each site was mapped. This was achieved by performing one or more double digests using the enzyme of interest, together with an enzyme cleaving the plasmid at a known position. Restriction fragments were separated by electrophoresis alongside Hind III/EcoRI-digested bacteriophage lambda DNA size markers on 1% (w/v) agarose, 1 x TAE gels pre-stained with ethidium bromide. The distance of migration of each plasmid DNA fragment was measured, as well as the migration distances of the size markers. The size of each plasmid DNA fragment was then determined by means of a computer programme written for the BBC model B microcomputer and kindly furnished by N. Crickmore. This data was then used to assemble complete restriction maps of each clone.

SECTION II.11. DIDEOXYNUCLEOTIDE SEQUENCING OF LECTIN CLONES

DNA sequencing was carried out using the dideoxynucleotide chain termination method of Sanger et al (1977).

Additional sequence data, kindly provided by R. Barker, was obtained by the chemical cleavage method of Maxam and Gilbert (1980).

II.11.A. Subcloning of fragments into M13 vectors

In order to produce single-stranded DNA templates for the dideoxy-sequencing reactions, it was first necessary to clone the DNA fragments of interest into M13 vectors. Double stranded replicative form (RF) DNA of the vectors mp18 and mp19 was digested and dephosphorylated as previously described (section II.8.A.). DNA fragments to be cloned into these vectors were derived from the plasmid subclones. Digestion of the plasmid DNAs and purification of the desired fragments was performed as described previously (section II.8.B.), as was the ligation of the purified DNA fragments to the dephosphorylated vector DNAs (section II.8.C.). The ligated DNAs were then used to transform competent cells made from the E. coli strain JM101. The same procedure as that described in section II.8.D was used, except that following heat shock, the cells were added to 3 ml of molten H-Top agar (1% (w/v)

bacto-tryptone, 136 mM NaCl, 0.8% (w/v) nutrient agar), previously melted and maintained at 45°C. Also added to this mixture was 25 μ l of 25 mg/ml IPTG, 25 μ l of 20 mg/ml X-gal and a 200 μ l aliquot from a culture of exponentially growing cells of the E. coli strain JM101. After gentle mixing, the cell suspension was plated out onto 1.5% (w/v) L-agar plates and the top agar allowed to set. The plates were then inverted and incubated at 37°C overnight.

II.11.B. Preparation of single-stranded DNA templates

The method used was taken from the M13 cloning and sequencing handbook, Amersham (1984). By means of a sterile toothpick, each plaque of interest was used to inoculate a 1.5 ml culture in a 25 ml McCartney bottle. The medium inoculated was obtained by diluting an overnight culture of E. coli strain JM101 grown in 2 x TY medium (1.6% (w/v) bacto-tryptone, 1% (w/v) yeast extract, 85 mM NaCl) by a factor of 100, again with 2 x TY. The 1.5 ml cultures were incubated at 37°C with vigorous shaking (300 rpm) for 5½ h. Each culture was then transferred to a 1.5 ml Eppendorf tube and centrifuged in a microcentrifuge for 5 min. The supernatants were poured off into fresh Eppendorf tubes and recentrifuged for a further 5 min. The resultant supernatants were again poured carefully into fresh tubes and 200 μ l of 20% (w/v) PEG 6000, 2.5 M NaCl was added to each. Following

mixing the tubes were left to stand at room temperature for 15 min. Bacteriophage precipitates were pelleted by centrifugation in a microcentrifuge for 2 min. The supernatants were removed by means of a micropipette and the tubes re-centrifuged briefly to bring any remaining liquid to the bottom of the tube. Final traces of the supernatants were removed, again using a micropipette, and the viral pellets were each resuspended in 100 μ l of TE and 50 μ l of TE-saturated phenol. The tubes were vortexed for 20 s and then incubated at room temperature for 20 min. Re-vortexing of the samples followed by centrifugation for 2 min was then carried out. The aqueous phases were each re-extracted with 50 μ l of 24:1 (v/v) chloroform/amy alcohol mixture and the DNA precipitated by the addition of 10 μ l of 3 M sodium acetate pH 6.0 and 250 μ l of ethanol to each tube. The preparations were stored at -20°C overnight and the DNA pelleted by centrifugation for 10 min in a microcentrifuge. The pellets were washed twice in 70% (v/v) aqueous ethanol, dried in vacuo and finally each redissolved in 20 μ l of sterile distilled water. The single-stranded DNA preparations were stored at -20°C .

II.11.C. Sequencing reactions

DNA sequencing was performed using [α - ^{35}S]dATP γ S as the radiolabelled nucleotide (Biggin et al, 1983). The method

from the Amersham M13 cloning and sequencing handbook, (1984) was followed. For each of the 4 bases, deoxynucleotide/dideoxynucleotide working solutions were made as outlined in Table 3, using the stock solutions supplied in the Amersham sequencing kit. Single-stranded template DNA (3ul) was annealed to 1 ul of 1.2 ug/ul M13 17-mer sequencing primer in 1.5 x Klenow reaction buffer (154 mM tris-HCl pH8.7, 5 mM MgCl₂) in a final volume of 10 ul for 2 h at 60°C in an oven. The annealed template was cooled on ice and 1 ul of [α -³⁵S]dATP<S (>400 uCi/mmol) was added. Klenow polymerase was diluted to 1u/ul in 1 x Klenow reaction buffer and 1 ul was added to the annealed template/radiolabel mix. A 2.5 ul aliquot from the template/radiolabel/Klenow mix was placed into each of four 1.5 ml Eppendorf tubes labelled G,A,T and C, each containing 2 ul of the appropriate nucleotide working solution. Gentle mixing was achieved by means of a micropipette and the samples were then incubated at 30°C for 20 min. 2 ul of chase solution (0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dGTP, 0.5 mM dTTP) was added to each tube and after gentle mixing, the samples were incubated for a further 15 min at 30°C. To each sample was then added 4 ul of formamide dye mix (0.1% (w/v) xylene cyanol, 0.1% (w/v) bromophenol blue, 20 mM EDTA in deionised formamide, prepared as before). If not required for immediate use, sequencing samples were stored at -20°C.

Table 3

	A	C	G	T
0.5mM dCTP	20 ul	1 ul	20 ul	20 ul
0.5mM dGTP	20 ul	20 ul	1 ul	20 ul
0.5 mM dTTP	20 ul	20 ul	20 ul	20 ul
1 x TE buffer	20 ul	20 ul	20 ul	20 ul
0.15 mM ddATP	80 ul	-	-	-
0.02 mM ddCTP	-	61 ul	-	-
0.05 mM ddGTP	-	-	61 ul	-
0.5 mM ddTTP	-	-	-	80 ul

Table 3 Composition of nucleotide working solutions used
for M13 dideoxy-sequencing

The four nucleotide working solutions required for M13 dideoxy-sequencing were prepared from 0.5 mM dNTP and 0.5 mM ddNTP stocks as described.

II.11.D. Polyacrylamide gel electrophoresis of sequencing products

Electrophoresis was carried out using salt gradient polyacrylamide gels. This enables longer sequences to be read from a single gel (Biggin et al., 1983). A filter sterilised stock solution of 38% (w/v) acrylamide, 2% (w/v) NN'-methylenebisacrylamide was used to prepare working solutions for the preparation of the gradient gels. These were respectively 0.5 x TBE, 5.7% (w/v) acrylamide, 0.3% (w/v) NN'-methylenebisacrylamide, 7 M urea and 2.5 x TBE, 5.7% (w/v) acrylamide, 0.3% (w/v) NN'-methylenebisacrylamide, 7 M urea, 5% (w/v) sucrose, 50 ug/ml bromophenol blue. Each sequencing gel was prepared from 22 ml of the 0.5 x TBE mixture and 8 ml of the 2.5 x TBE mixture. Polymerisation was achieved by the addition to each solution of 6 ul of 10% (w/v) ammonium persulphate per ml and 0.5 ul of TEMED per ml. The salt gradients were prepared by drawing first the 0.5 x TBE mixture and then the 2.5 x TBE mixture into a 25 ml pipette. A little air was also drawn in to achieve mixing. The gels were immediately poured. 40 cm glass gel plates were used in conjunction with BRL 0.4 mm spacers and a BRL sharktooth comb producing 5.7 mm wide wells. Polymerisation was allowed to proceed for 1 h prior to assembly of the gel apparatus. The buffers used in the top and bottom reservoirs were 0.5 x TBE and 2.5 x TBE respectively.

Prior to use, gels were pre-electrophoresed at 40 W for 15 min and the wells cleared of salt and unpolymerised acrylamide using a micropipette. Sequencing samples were heated at 95°C for 3 min prior to loading.

Electrophoresis was carried out at 40W until the bromophenol blue dye front had reached the end of the gel. This normally took approximately 100 min. After the gel spacers and comb had been removed, the glass plates were separated using a single-edged razor blade. The plate to which the gel had adhered was placed in 10% (v/v) methanol, 10% (v/v) acetic acid for 15 min in order to fix the DNA. The gel was then transferred onto Whatman 3MM paper, dried in vacuo at 80°C for 50 min and finally autoradiographed overnight at room temperature using X-omat S X-ray film.

SECTION 11.12. ISOLATION OF RNA FROM CASTOR BEAN TISSUES

11.12.A. Isolation of total RNA from castor bean endosperm

Total RNA was extracted from castor bean endosperm tissue using a modification of the method described by Keller and Taylor (1976). In the case of tissue harvested from germinating seeds, the embryo and testa were removed prior to homogenisation. However, in the case of endosperm harvested from developing or dry seeds, no attempt was made to excise the relatively minute embryos. Thus only the testas, where present, were removed. The harvested tissue, selected for uniformity at the appropriate developmental stage, was frozen in liquid nitrogen, ground to a fine powder and homogenised with 3 ml of extraction buffer (50 mM tris-HCl pH 9.0, 150 mM NaCl, 5 mM EDTA, 5% (w/v) SDS) per g of material using a mortar and pestle. The slurry was added to an equal volume of TE-aquilibrated 1:1 (v/v) redistilled phenol/chloroform mixture and mixed for 10 min at room temperature. The phases were then separated by centrifugation and the aqueous phase removed and saved. After 2 further phenol/chloroform extractions had been carried out, 5 M NaCl was added to give a final concentration of 0.2 M. The nucleic acids were then precipitated overnight at -20°C by the addition of 2.5 volumes of ethanol. The precipitated nucleic acids were

collected by centrifugation at 15 000 g, 2°C for 15 min and washed 4 times in 3 M sodium acetate pH 5.5. The acetate washed pellet was redissolved in a small volume of 0.3 M NaCl as before. After any insoluble debris had been removed by centrifugation at 15 000 g and 2°C for 10 min, the nucleic acids were precipitated with 2.5 volumes of ethanol as before. Finally the precipitate was collected by centrifugation at 12 000 g and 2°C for 30 min, washed 3 times in 70% (v/v) aqueous ethanol, dried in vacuo and dissolved in a small volume of sterile distilled water. The A₂₆₀ of a 2 ul aliquot of each preparation was determined in order to quantify the yield of RNA (Maniatis et al, 1982)

II.12.B. Isolation of total RNA from castor bean roots and leaves

When the method of RNA extraction previously described was used on castor bean root and leaf tissue, it was found that negligible yields of RNA were obtained. RNA extraction was therefore attempted using a modification of the method described by Logemann et al (1986). Again only very small quantities of total RNA (approximately 1 ug per g of tissue) were obtained. Although such yields are unacceptably low, there was just sufficient RNA obtained to perform the experiments in this study. It appeared, from gel electrophoresis of the samples, that some RNA

degradation may have taken place. Although this should be borne in mind when interpreting the results obtained, the samples were nevertheless considered to be of sufficient quality to enable the detection of specific transcripts. Initially the plant tissue was frozen in liquid nitrogen and ground using a mortar and pestle. 2 ml of an extraction buffer (8 M guanidinium-HCl, 20 mM 4-morpholine ethanesulphonic acid, 20 mM EDTA, 50 mM 2-mercaptoethanol made to pH 7.0 with NaOH) was added per g of tissue once thawing had started and further homogenisation was carried out. An equal volume of TE-equilibrated 1:1 (v/v) redistilled phenol/chloroform mixture was then added and the preparation thoroughly mixed. The phases were separated by centrifugation at 5 000 g for 10 min and the aqueous phase removed. Another extraction with phenol/chloroform was carried out in the same way and the nucleic acids were precipitated by the addition of a tenth volume of 1 M acetic acid and an equal volume of ethanol, prior to storage at -20°C overnight. The precipitate was collected by centrifugation at 15 000 g and 2°C for 15 min and washed 3 times with ice-cold 3 M sodium acetate pH 5.5. Finally, the pellet was washed once with 75% (v/v) aqueous ethanol, gently dried in vacuo and redissolved in a small volume of sterile distilled water. RNA yields were quantified as before.

SECTION II.13. NORTHERN BLOTTING AND HYBRIDISATION

II.13.A. RNA gels

RNA samples were electrophoresed on 100 ml gels containing 50% (v/v) formamide (deionised as before), 1.2% (w/v) agarose and 0.1 x TEP (3.6 mM tris, 3 mM NaH_2PO_4 , 2 mM EDTA made to pH 8.0 with HCl). Leicester Biocentre 10 cm gel tanks were used for this purpose. 10 μg aliquots from the RNA preparations, contained in a volume of 4 μl or less, were each mixed with 20 μl of loading buffer (60% deionised formamide, 0.1 x TEP). The samples were then heated at 65°C for 5 min and placed on ice. Following the addition to each of 3 μl of 50% (v/v) glycerol, 0.25% (w/v) bromophenol blue, the samples were electrophoresed at 20 mA for 1-2 hours in 0.1 x TEP gel buffer covering the sides but not submerging the gel.

II.13.B. Northern transfer and hybridisation

Prior to transfer onto Hybond-N membrane, RNA gels were soaked for 1 h in 20 x SSC. In all other respects the procedure used for Northern transfer was identical to that used for the transfer of DNA onto nylon membrane (section 4.C.). Hybridisation was carried out using a modification of the method detailed in the Amersham membrane transfer and detection methods handbook. Membranes were pre-

hybridised and hybridised in 5 x SSPE (0.9 M NaCl, 50 mM NaH_2PO_4 , 0.5 mM EDTA), 50% (v/v) deionised formamide, 5 x Denhardt's solution, 0.5% (w/v) SDS, 100 ug/ml polyadenylic acid at 42°C for at least 14 hours. All other details were as described in section II.4.D., the membranes being washed under low stringency conditions.

SECTION II.14. ANALYSIS OF IN VIVO TRANSCRIPTION OF
LECTIN GENES USING RNASE PROTECTION

II.14.A. Construction and linearisation of plasmids for
transcription

In order to generate gene-specific probes, restriction fragments were excised from the plasmid subclones and cloned into the plasmid transcription vectors pGEM-blue3 and pGEM-blue4 (see appendix I). Each construction was made so as to allow the generation of transcripts antisense to the in vivo mRNAs using T7 RNA polymerase. DNA manipulations were carried out as previously described and the recombinant plasmids were each linearised at an appropriate polylinker site at the end of the DNA insert distal to the T7 promoter. The restriction digests were extracted once with TE-equilibrated 1:1 (v/v) phenol/chloroform mixture and the DNA precipitated by the addition of a tenth volume of 3 M sodium acetate pH 4.8 and 2 volumes of ethanol. After the precipitate had been recovered by centrifugation for 10 min in a microfuge and dried in vacuo, the linearised DNA was redissolved in TE at a concentration of 0.5 ug/ul.

II.14.B. Preparation of radiolabelled RNA probes

The methods described by Kreig and Melton (1986) were used for both the in vitro synthesis of radiolabelled RNA probes and the subsequent RNase protection experiments performed. Transcription reactions were assembled at room temperature by mixing the following:- 1.5 ul of sterile distilled water, 1 ul of 0.5 ug/ul linearised DNA, 1 ul of 200 mM dithiothreitol, 0.5 ul of human placental RNase inhibitor, 1 ul of 1 mg/ml nuclease-free bovine serum albumin, 1 ul of CGA ribonucleotide mix (5 mM GTP, 5 mM GTP, 5 mM ATP), 1 ul of 100 uM UTP, 1 ul of ^{32}P -UTP (20 uCi, >800 Ci/mmol), 1ul of 10 x transcription buffer (0.4 M tris-HCl pH 7.5, 60 mM MgCl_2 , 20 mM spermidine-HCl) and 1 ul of T7 RNA polymerase. After the mixture had been incubated for 40 min at 37°C, 0.5 ul of RNase-free DNase (23 u/ul) was added and the preparation incubated for a further 10 min at room temperature. To the preparation was then added 40 ul of sterile distilled water, 25 ul of TE-equilibrated redistilled phenol and 25 ul of chloroform. After thorough vortexing and centrifugation for 2 min in a microfuge, 50 ul of the aqueous phase was removed into a fresh tube. To this tube was added 5 ul of 7 M ammonium acetate, 1 ul of 1 ug/ul tRNA and 150 ul of ethanol. The preparation was placed on dry ice for 30 min and the RNA precipitate collected by centrifugation in a microfuge for 10 min. The pellet was then dried in vacuo, redissolved in 5 ul of formamide sequencing dyes and heated at 95°C for 3 min. The sample was loaded onto a

standard 6% (w/v) acrylamide sequencing gel which was electrophoresed at 40 W for 1 h. Following the removal of one of the glass plates, the gel was covered with cling film and 3 stickers were positioned around its edges. Distinguishing marks were made on each of these stickers using ink containing ^{32}P -radiolabelled nucleotides (2 uCi/ul) and the gel was exposed to Kodak X-omat S X-ray film for 2 min. After the film had been developed and dried, alignment with the gel was achieved by means of the distinguishing marks and the position of the full-length transcription product was marked on the gel using a heated needle. The transcript band was then excised and transferred to an Eppendorf tube. The radiolabelled RNA was eluted by incubation for 1 h at 60°C in 200 ul of elution buffer (0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% (w/v) SDS). After the eluate had been transferred to a fresh tube, 500 ul of ethanol was added and the tube thoroughly vortexed. A 10 ul aliquot was then removed and its activity determined in a scintillation counter.

II.14.C. RNAse protection

Aliquots containing 10 ug of RNA were removed from castor bean RNA samples which had previously been ethanol precipitated and thoroughly vortexed. Each was mixed in a 1.5 ml Eppendorf tube with an aliquot of the ethanol

precipitated RNA probe containing 50 000 cpm of radioactivity. In addition two control mixtures were assembled, each containing 10 ug of ethanol precipitated E. coli tRNA and an aliquot of the RNA probe as before. The samples were centrifuged for 10 min in a microcentrifuge. Each RNA pellet was then dried in vacuo, resuspended in 24 ul of deionised formamide and mixed with 6 ul of 5 x hybridisation buffer (2 M NaCl, 200 mM PIPES pH 6.4, 5 mM EDTA). The samples were heated at 85°C for 4 min and incubated overnight at 45°C. To each tube, apart from one of the two control tubes, was added 150 ul of RNase digestion solution (300 mM NaCl, 10 mM tris-HCl pH 7.5, 5 mM EDTA, 40 ug/ml ribonuclease A, 2 ug/ml RNase T1). To the remaining tube was added 150 ul of a similar solution lacking RNase but otherwise of identical composition to the digestion solution. After mixing, the samples were incubated for 30 min at 37°C. To each tube was then added 2 ul of freshly prepared 25 mg/ml proteinase K and 10 ul of 10% (w/v) SDS and the samples were incubated at 37°C for a further 15 min. The samples were then each extracted with 200 ul of TE-equilibrated 1:1 (v/v) phenol/chloroform mixture. The phases were separated by centrifugation for 2 min in a microcentrifuge and 160 ul of each aqueous phase was removed into a fresh 0.5 ml Eppendorf tube. To each sample was then added 1 ul of 5 ug/ul E. coli tRNA, 8 ul of 7 M ammonium acetate and 320 ul of ethanol. After the tubes had been chilled on

dry ice for 30 min, the precipitated RNA was collected by centrifugation in a microcentrifuge for 20 min. Most of each supernatant was removed so as to leave a volume of approximately 20 μ l of liquid in each tube. The remaining supernatant was then removed by incubating the tubes in a rotary vacuum desiccator for 40 min. The RNA pellets were each redissolved in 5 μ l of formamide sequencing dyes, with the exception of the non-RNase treated sample, which was resuspended in 20 μ l of the same dye mixture. Also prepared for electrophoresis in a similar way were 32 P-radiolabelled, Hpa II-digested pBR322 DNA size markers. Finally, the samples were heated to 95°C for 3 min and centrifuged for 1 s in a microcentrifuge to ensure that all liquid was collected at the bottom of each tube. A 3.5 μ l aliquot was removed from each sample and loaded onto a standard 6% (w/v) acrylamide sequencing gel, prepared as previously described except that a standard type 0.4 mm thickness home-made gel comb producing 5 mm wide wells was used rather than a sharktooth comb. The gel was electrophoresed, fixed, dried and autoradiographed as described in section II.12.D.).

SECTION II.15. RADIOACTIVE LABELLING OF DNA IN VITROII.15.A. Nick translation of DNA fragments

Purified DNA restriction fragments, isolated as described in section II.8.B., were labelled by nick translation using a modification of the method of Rigby et al (1977). Initially, a 30 μ l reaction mixture was assembled containing the following components:- 100 ng of purified DNA, 3 μ l of 10 x nick translation buffer (0.5 M tris-HCl pH 7.8, 50 mM $MgCl_2$, 100 mM dithiothreitol), 1 μ l of 100 μ M dATP, 1 μ l of 100 μ M dTTP, 1 μ l of 100 μ M dCTP, 1 μ l of 1 μ M dGTP and 1-6 μ l of (α - ^{32}P)dGTP (>3 000 Ci/mmol), (the amount depending on the required specific activity of the probe). To this mixture was added 1 μ l of freshly prepared 50 μ g/ml DNase I (from bovine pancreas). Following incubation for 15 min at room temperature, 1 μ l of E. coli DNA polymerase I (Kornberg) was added. The preparation was then gently mixed and incubated for 3 h at 15°C. Removal of the unincorporated nucleotides was achieved by spermine precipitation. To this end, 3 μ l of 100 mM spermine-HCl and 5 μ l of 1 μ g/ μ l E. coli tRNA were added to the preparation which was then lightly vortexed. The tube was placed on ice for 15 min and then centrifuged for 10 min in a microcentrifuge to pellet the nucleic acids. The pellet was washed twice with 100 μ l of 70% (v/v) ethanol, 0.3 M sodium acetate pH 8.5, 1.25 mM EDTA,

twice with 100 μ l of 70% (v/v) aqueous ethanol, dried in vacuo and redissolved in 100 μ l of sterile distilled water. This nick translation method was used to prepare DNA probes for the screening of the bacteriophage lambda genomic libraries and most of the Southern blots presented.

II.15.B. Oligolabelling of DNA fragments

A modification of the oligolabelling method of Feinberg and Vogelstein (1984) was used to prepare ^{32}P -radiolabelled DNA probes of high specific activity. Such probes were used for hybridisation to both Northern blots and some genomic Southern blots. Initially a preparation of the DNA fragment of interest, purified as described in section II.8.B., was denatured by incubation for 5 min in a boiling water bath. The preparation was then transferred to a water bath at 37°C for 5-10 min. Meanwhile the remainder of the reaction mixture was assembled in a 1.5 ml Eppendorf tube. The mixture consisted of 3 μ l of 5 x oligolabelling buffer (0.24 M tris-HCl pH 8.0, 25 mM MgCl_2 , 50 mM 2-mercaptoethanol, 0.1 mM dATP, 0.1 mM dTTP, 0.1 mM dCTP, 1 M HEPES pH 6.6, 54 A₂₆₀ units/ml hexadeoxyribonucleotides (from calf thymus DNA), 0.6 μ l of 10 mg/ml nuclease-free bovine serum albumin, 2-5 μ l of (α - ^{32}P)dGTP ($>3\ 000\ \text{Ci/mmol}$) and 0.6 μ l of 3.5 u/ μ l DNA polymerase I Klenow fragment. A 10 ng

aliquot of the DNA preparation was added to this mixture and the final volume was adjusted to 15 ul by the addition of sterile distilled water. The preparation was incubated at room temperature for a minimum of 4 h and the unincorporated nucleotides removed by spermine precipitation as described in section II.16.A.

II.15.C. Radiolabelling of DNA restriction fragments using an end filling reaction

Hind III-digested bacteriophage lambda DNA fragments and Hpa II-digested pBR322 plasmid DNA fragments were radiolabelled with (α -³²P)dGTP and (α -³²P)dCTP respectively using a modification of the method of Drouin (1980). The bacteriophage or plasmid DNA was first digested with the appropriate restriction endonuclease at a concentration of 0.05 ug/ul. To 20 ul of the digested DNA preparation was added 1 ul of 3.5 u/ul DNA polymerase I Klenow fragment and either 2 ul of (α -³²P)dGTP (>3 000 Ci/mmol) in the case of Hind III-digested bacteriophage lambda DNA, or 2 ul of (α -³²P) dCTP (>3000 Ci/mmol) in the case of Hpa II-digested pBR322 plasmid DNA. The preparation was then incubated at room temperature for 20 min and extracted once with an equal volume of TE-equilibrated 1:1 (v/v) phenol/chloroform mixture. The DNA was precipitated by the addition of 2.5 ul of 7 M ammonium acetate and 50 ul of ethanol. The precipitate was then

collected by centrifugation for 10 min in a microcentrifuge, dried in vacuo and redissolved in 300 ul of TE.

SECTION III

RESULTS

SECTION III.1. GENOMIC ORGANISATION OF CASTOR BEAN LECTIN GENES

III.1.A. Estimation of *R. communis* genome size

It is helpful to know the genome size when attempting to clone specific genes from an organism, since an estimate may be made as to how many recombinant clones must be screened. Unfortunately, the C-value or haploid genome size of *R. communis* is not available in the literature. An approximate estimate was therefore made by comparing the DNA yields obtained from known numbers of nuclei. Table 4 shows the respective DNA yields and nuclei counts of a number of preparations, as well as the individual and mean C-values calculated. The results shown are discussed in section IV.1.C.1).

II.1.B. Estimated number of lectin genes in the *R. communis* genome

In order to estimate the size of the lectin gene family of *R. communis*, a genomic Southern blot was performed as described in Figure 5. The blot was hybridised with a ricin cDNA probe and washed at low stringency, so that both ricin and RCA I-like genes would be identified in the resulting autoradiograph. The result, shown in Figure 5

Table 4

Sample	No of nuclei	DNA yield	Calculated C-value
1	2.1×10^8	180 ug	0.43 pg
2	5.4×10^8	335 ug	0.31 pg
3	5.7×10^8	215 ug	0.19 pg
4	1.4×10^9	650 ug	0.23 pg

Mean C-value = 0.26 pg

Confidence limits (95% level) = + or - 0.17 pg

Table 4 Estimation of *R. communis* C-value from nuclei
counts and DNA yields

In order to estimate the *R. communis* haploid genome size, the DNA yields obtained from preparations containing known numbers of nuclei were quantified. A C-value was calculated in each case and an overall mean value, plus confidence limits, was determined. Although the *R. communis* seedlings used were of unknown variety, it was assumed that the cultivar used is diploid, as has been shown for other cultivars (D Griffiths, pers. comm.).

The results shown are discussed in Section IV.1.C.1).

Figure 5

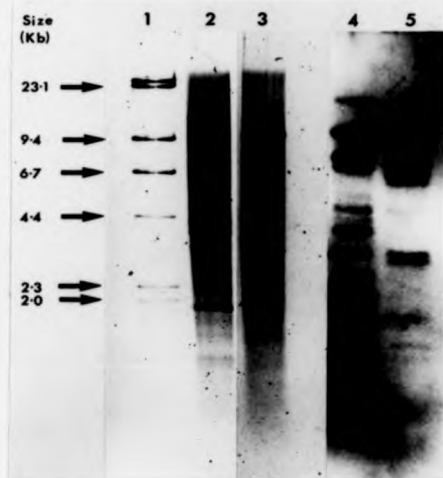


Figure 5 Identification of genomic DNA restriction
fragments containing lectin-hybridising
sequences

Two 10 ug samples of castor bean nuclear DNA were digested with restriction enzymes and electrophoresed on a 1 x TAE, 0.6% (w/v) agarose gel. The gel was photographed under uv illumination following ethidium bromide staining. A polaroid negative of the gel is shown in tracks 1 to 3. The size markers in track 1 are of Hind III - digested bacteriophage lambda DNA. Tracks 2 and 3 show the patterns obtained when castor bean nuclear DNA is digested with Eco RI and Hind III respectively.

The gel was Southern blotted onto Hybond-N and the membrane was probed with a nick-translated ricin cDNA fragment (sections II.4.C., II.15.A.). After low stringency washing (section II.4.D.), the membrane was autoradiographed. Tracks 4 and 5 show the size distribution of lectin-hybridising fragments in castor bean DNA digested with Eco RI and Hind III respectively.

indicates that the R. communis lectin gene family consists of approximately 8 members. A detailed discussion is given in section IV.4.A.

In order to estimate how many ricin-like genes are present in the R. communis genome, a second genomic blot was probed with the same ricin cDNA fragment, but washed instead at high stringency, as described in Figure 6. The result, which is also discussed in section IV.4.A., appears to indicate that two members of the lectin gene family are ricin-like.

Figure 6

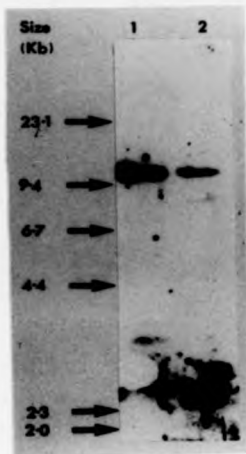


Figure 6 Identification of genomic DNA restriction
fragments containing sequences sharing
high homology with a ricin cDNA probe

Two 10 ug samples of castor bean nuclear DNA were digested with restriction enzymes and electrophoresed on a 1 x TAE, 0.6% (w/v) agarose gel alongside Hind III - digested lambda DNA size markers. The gel was Southern blotted onto nitrocellulose and probed with a nick-translated ricin cDNA probe (sections II.4.C., II.15.A.). A high stringency wash was then carried out (section II.4.D.) so as to specifically identify DNA fragments containing ricin-like sequences. The position of the size markers is indicated. Tracks 1 and 2 show the patterns of hybridisation obtained from Hind III and Eco RI-digested castor bean DNA respectively.

SECTION III.2. NUMBER AND DIVERSITY OF CLONED LECTIN
GENES

III.2.A. Construction and screening of genomic libraries

Two R. communis genomic libraries were constructed during the course of this study, one from partial Eco RI digests and one from partial Sau 3A digests. Apart from the restriction enzymes involved, the procedure used to prepare the two libraries was identical. The overall cloning strategy used to prepare the partial Sau 3A library is shown in Figure 4. The lambda vector charon 35 (Loenen and Blattner, 1983) was used for the construction of both genomic libraries. In addition to the usual features of bacteriophage lambda replacement vectors, the charon 35 vector has two additional features which made it particularly useful in this study. These are as follows:-

- 1) A polylinker sequence containing seven unique restriction sites is present on either side of the stuffer fragment. Genomic libraries may therefore be constructed from various types of restriction digest and more importantly, if the Bam HI cloning sites are used, the remaining sites of the polylinkers are retained. Thus even though one or both of the Bam HI sites are often destroyed when Sau 3A fragments are cloned into the

vector, the six remaining sites may still be used to cleanly excise any cloned inserts.

ii) The charon 35 vector has been specifically designed for use with recombination-deficient host strains. This was achieved by engineering the lambda gam gene situated in the stuffer region of earlier vectors into the right arm of the new vector. Any recombinants produced will therefore be gam⁺, thus allowing their propagation on E. coli rec A⁻ strains. Furthermore, the gam gene product inhibits the exonuclease V activity of the E. coli rec B,C genes, so charon 35 recombinants are phenotypically rec B,C⁻ as well as rec A⁻. This provides a double block to recombination. Constructions of this type have been shown to be particularly useful for the preparation of genomic libraries from higher plants such as wheat (Murray et al, 1984), where large quantities of interspersed repetitive DNA are present in the genome. These repetitive sequences may well be responsible for the insert instability problems associated with some lambda vectors (Murray et al, 1984), since they frequently contain internal homologies.

A diagram illustrating the essential features of the charon 35 vector is shown in the Appendix I. Details of the experimental procedure followed for making the genomic libraries may be found in section II.5.

Initially, a number of partial restriction digests were prepared using progressively smaller quantities of restriction enzyme obtained from a dilution series. The products of the digests were electrophoresed on a 0.6% (w/v) agarose gel to separate the restriction fragments. Thus the optimum conditions for producing 10-22 Kb sized fragments could be identified. Figure 7 shows a gel containing the test digests carried out for the construction of the partial Sau 3A library. It can be seen by comparing the samples in tracks 1 and 2 that very little mechanical shearing has resulted from the mixing process. Virtually no noticeable digestion has occurred at 0.015 u/ug (track 3), but a progressively lower average size is observed amongst the DNA fragments in each successive sample. Complete digestion appears to have occurred at the enzyme/DNA ratio of 1.0 u/ug (track 9) and above. The sample which shows the greatest intensity of staining in the desired size range (10-22 Kb) is in track 5 (0.06 u/ug). Thus the same conditions were used to prepare partially digested DNA on a large scale for cloning. Large scale digests were also prepared using the two enzyme/DNA ratios immediately above and below the optimum. The digestion products obtained using all three conditions were pooled prior to size fractionation in order to maximise the randomness of the library obtained. Fractionation was achieved on 10%-40% (w/v) sucrose gradients. Selected fractions from the gradients were

Figure 7

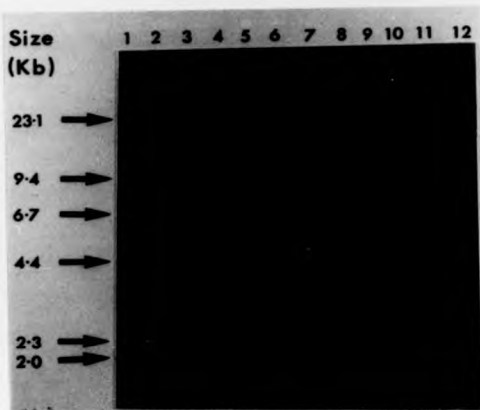


Figure 7 Assessment of optimal conditions for
partial digestion of castor bean DNA

2 ug aliquots of castor bean DNA were digested at a range of Sau 3A restriction enzyme concentrations. The procedure used to obtain the enzyme dilution series is described in section II.5.A. After digestion, the samples were electrophoresed alongside undigested castor bean DNA and Hind III-digested lambda DNA markers on a 0.5% (w/v) agarose, 1 x TAE gel. The position of the size markers is indicated by the arrows. Track 1 contains 2 ug of undigested castor bean DNA. Track 2 shows a second undigested sample which was subjected to the same mixing process as the enzyme-treated sample. Tracks 3 to 12 show the DNA samples which were treated with progressively larger amounts of Sau 3A. The exact enzyme/DNA ratios used were as follows:-

track 3, 0.015 u/ug; track 4, 0.03 u/ug; track 5, 0.06 u/ug; track 6, 0.12 u/ug; track 7, 0.25 u/ug; track 8, 0.5 u/ug; track 9, 1.0 u/ug; track 10, 2.0 u/ug; track 11, 4.0 u/ug; track 12, 8.0 u/ug.

size analysed by gel electrophoresis so as to identify fractions consisting of fragments in the size range 10-22 Kb. Figure 8 shows a typical gel obtained in this way. It appears that a good fractionation has been achieved, as judged by the smooth curve produced by the progressively smaller sized DNA samples. Fractions 16 to 22 were later pooled and aliquots were ligated with appropriately prepared arms of the lambda vector charon 35 in the stoichiometries detailed in Table 5. In the same table are detailed the yields of bacteriophages obtained from the respective ligated DNA samples. As can be seen from the table, the cloning efficiencies obtained in the construction of the partial Sau 3A library were far greater than that obtained in the construction of the partial Eco RI library. The most likely explanation for the low cloning efficiency obtained in the latter case was that the packaging mixture had been stored for some time and had therefore probably deteriorated. The results shown in Table 5 are discussed in detail in section IV.1.C.ii).

Two lectin-positive clones were identified amongst the 1.2×10^5 bacteriophages in the partial Eco RI library. Unfortunately, one of these clones proved impossible to purify. Restriction mapping and DNA sequencing analysis of the other clone, subsequently referred to as lambda clone 1, showed that it contains a lectin pseudogene

Figure 8

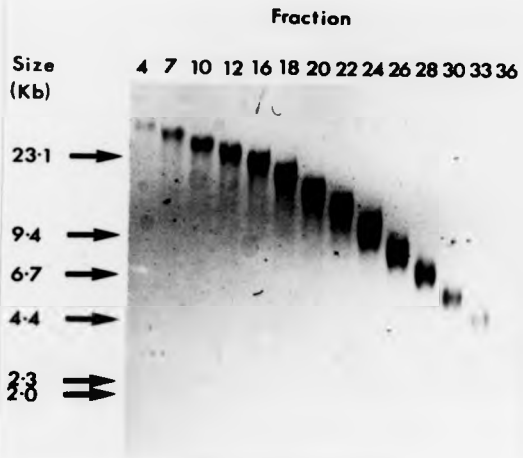


Figure 8 Size analysis of digested genomic DNA
fractions removed from a sucrose gradient

The partially digested genomic DNA to be used for cloning was size fractionated on a 10%-40% (w/v) sucrose gradient. After centrifugation, the tube containing the sucrose gradient was punctured at the bottom and successive 0.5 ml fractions were removed. 30 μ l aliquots from selected fractions, as indicated, were used for size analysis on a 0.4% (w/v) agarose, 1 x TAE gel. The samples were electrophoresed alongside a Hind III-digested lambda DNA sample containing NaCl and sucrose at the same concentration as occurs in the middle of the gradient. The position of the size markers is indicated by the arrows.

Table 3

Reaction	Restriction enzyme used to prepare genomic fragments	Quantity of genomic fragments used in ligation (ug)	Restriction enzyme used to prepare lambda arms	Quantity of lambda arms used in ligation (ug)	Yield of bacteriophages (pfu)	Cloning efficiency (pfu/ug)
1	<u>Eco</u> RI	0.33	<u>Eco</u> RI	0.67	1.2×10^5	1.2×10^5
2	<u>Sau</u> 3A	0.66	<u>Bam</u> HI	1.34	1.0×10^7	3.0×10^6
3	<u>Sau</u> 3A	1.0	<u>Bam</u> HI	1.0	3.0×10^6	2.5×10^6

Table 5 Yields of bacteriophages from lambda in vitro
packaging reactions

For the purposes of this study, two different castor bean genomic libraries were prepared; one from partial Eco RI digests and the other from partial Sau 3A digests. The two libraries were prepared at different times, but packaging reactions 2 and 3 were carried out in parallel using the same substrate DNA preparations and later pooled. Table 5 shows the yield of bacteriophages obtained from each packaging reaction and the corresponding cloning efficiencies calculated in each case.

The results shown are discussed in section IV.1.C.ii).

(Figures 17.b and 17.c). Thus, since the partial Eco RI library yielded no clones containing functional genes, the second genomic library was constructed from partial Sau 3A digests. Since Sau 3A recognises a tetranucleotide sequence, in contrast with Eco RI, which recognises a hexanucleotide sequence, it was anticipated that the second library would contain a more random distribution of genomic clones. Moreover, a much larger library of some 1.5×10^7 bacteriophages was obtained, as can be seen in Table 5. 9×10^5 unamplified bacteriophages from the partial Sau 3A library were screened at low stringency with a ricin cDNA probe (section II.6.). A large number of potential lectin-positive plaques were identified from which 24 were selected for rescreening. This number probably represented only a third to a half of the total number of potential positive plaques. Care was taken to pick a range of plaques which produced a variety of signal intensities in the first round of screening. This precaution was taken to minimise the possibility that certain genes sharing greater homology with the probe might be selected in preference to those which were less homologous. Nevertheless, it was assumed that differential hybridisation would not be occurring to any great extent at the stringency used. This was inferred from the result of a genomic blot where an RCA I cDNA was used as a probe instead of a ricin cDNA. The pattern of lectin-homologous bands obtained with this blot (for which

nylon was used rather than nitrocellulose) was essentially the same as that obtained with the ricin cDNA probe.

Of the 24 putative positives selected from the first screening of the partial Sau 3A library, 16 clones (subsequently known as clones 2-17) hybridised positively in the second round of screening. Each of these 16 lectin positive clones was plaque purified. Thus, a total of 17 lectin positive clones were obtained from the two R. communis genomic libraries constructed. It should be noted that lambda clone 1 was analysed before the second genomic library was constructed. Nevertheless the characterisation of all the lambda clones will be discussed in parallel.

III.2.B. Southern analysis of lectin-positive lambda clones

DNA minipreparations were made from each of the lectin-positive lambda clones. Two aliquots were removed from each preparation (apart from that of lambda clone 1). One sample was digested with Eco RI and one with Hind III. No Hind III digest was performed with the DNA from lambda clone 1, as the Hind III polylinker sites of the vector were removed during the preparation of the partial Eco RI library. The digests were electrophoresed alongside DNA size markers on two 0.7% (w/v) agarose gels which were

subsequently blotted onto nitrocellulose. The membranes were hybridised with a ricin cDNA probe and washed at low stringency prior to autoradiography. The autoradiographs of the Southern blots obtained from the two gels are shown in Figures 9 and 10. On the basis of the pattern and intensities of the bands seen in each track, the clones were classified into five groups, A to E. The classification and the data upon which it is based is detailed in Table 6. Since the lectin pseudogene in lambda clone 1 had already been characterised, the identification of similar clones was relatively straightforward. Some of the groups may well contain clones of more than one gene, but the classification was considered sufficient for the purpose of selecting specific lambda clones for further analysis.

Figure 9

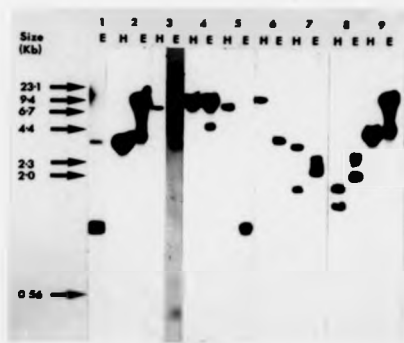


Figure 10



Figures 9 and 10 Southern blotting analysis of lectin-
positive lambda clones

DNA minipreparations were made from each lectin-positive lambda clone. Two 1 ug aliquots were removed from each preparation and separately digested with Hind III and Eco RI. The digests were electrophoresed on 0.7% (w/v) agarose, 1 x TAE gels in the presence of Hind III digested lambda DNA markers. After electrophoresis, the gels were Southern blotted onto nitrocellulose and the membranes probed with a nick translated ricin cDNA fragment (sections II.4.C., II.15.A.). The membranes were washed at low stringency (section II.4.D.) and autoradiographed overnight. The hybridisation patterns obtained are shown. Hind III and Eco RI digests are denoted respectively by the letters H and E, whilst the numbers above these letters identify the lambda clones. Two plasmid digests were included in this experiment as controls. Track C1 was loaded with a Pst I digest of the ricin cDNA-containing plasmid pRCL611, whilst track C2 was loaded with a Pst I digest of the RCA I cDNA-containing plasmid pRCL521.

Table 6

Group	Classes	No of hybridising Eco RI fragments observed	No of hybridising Hind III fragments observed	Subclone(s)	Relative strength of hybridisation signal (to ricin cDNA)
A	1, 5, 17	2 ^{oo}	Not applicable	pCBG1B2, pCBG1B3	Weak
B	2	3	1	pCBG2W1	Weak
C	2, 4, 6 ^a , 6, 10, 11, 12 ^a , 13, 15, 16 ^a	1	1	pCBG3B1	Strong
D	7, 8	3	2	pCBG4W2, pCBG4B4	Weak
E	14	2	2	-	Weak

The data above are discussed on page 193

Table 6 Classification of lambda lectin positive
clones and nomenclature of subclones in pUC 8

The 17 lectin-positive lambda clones were classified into five groups as shown, on the basis of the Southern blots shown in Figures 9 and 10. The clones chosen for further analysis are indicated by underlining, whilst the corresponding subclones obtained in each case are listed. It should be noted that some of the clones classified in group C, which have been marked with an asterisk, differed from the rest of the group in that they gave weaker hybridisation signals. This was apparently caused however, by poor DNA yields in the lambda minipreparations. The double asterisk indicates in the case of the group A clones that although only 2 lectin homologous Eco RI fragments could be identified in each case, a third fragment with an expected size of 344 bp was also assumed to be present.

SECTION III.3. DETAILED RESTRICTION ANALYSIS OF SELECTED CLONES

In total, four different lectin-positive lambda clones were chosen for further analysis by restriction mapping and DNA sequencing, namely clones 1, 3, 7 and 10 (Table 6). Restriction fragments were individually subcloned into the plasmid pUC8. Table 6 shows the restriction fragments isolated and the nomenclature of the subclones. The lambda DNA preparations used for subcloning were obtained using the "midiprep" method, with the exception of lambda clone 1, for which a large scale preparation was used (see section II.7.A).

In the case of lambda clone 1, the five different Eco RI fragments of the genomic insert which were visible on an ethidium bromide stained agarose gel were separately subcloned into pUC8 to produce pCBG1R1, pCBG1R2, pCBG1R3, pCBG1R4 and pCBG1R5. The individual numbers assigned to each of these plasmids corresponds to the order of size of their respective inserts. It was later shown that pCBG1R2 and pCBG1R5 contain respectively the 5' and 3' portions of a putative lectin gene, as well as some flanking sequences. The two subclones appeared to lack, however, a 344 bp region which overlaps both the A chain, linker peptide and B chain sequences in the ricin and RCA I

cDNAs. It was therefore assumed that an additional Eco RI fragment must be present in the lambda clone 1 insert. The fragment was probably not visible on the Southern blot (Figure 9) for one or both of two reasons. Firstly, since the fragment is small, its detection is not easily achieved on a 0.7% (w/v) agarose gel. Secondly, the fragment may share relatively little homology with the ricin cDNA sequence. No attempt was made to subclone the proposed 344 bp fragment from lambda clone 1, since it was subsequently shown by DNA sequencing that the potential lectin gene is non-functional (Figures 17.b and 17.c). The restriction maps of pCBG1R2 and pCBG1R5 are shown in Figure 11.

The subcloning of the putative lectin genes from lambda clones 3 and 10 was considerably simpler, since both could be excised on a single restriction fragment. A 7.3 Kb Hind III fragment subcloned from lambda clone 3 into pUC8 to produce the plasmid pCBG2H1 (Figure 12) was found to contain a putative lectin gene plus approximately 4.4 Kb of 5' flanking region and 1.1 Kb of 3' flanking region. Similarly, the 4.5 Kb Hind III fragment subcloned into pUC8 from lambda clone 10 to produce the plasmid pCBG3H1 (Figure 13) was found to contain a putative lectin gene with an identical restriction pattern to that of the ricin cDNA. The pCBG3H1 clone also contains approximately 2.2 Kb of 5' flanking region and 160 bp of 3' flanking region.

Figure 11

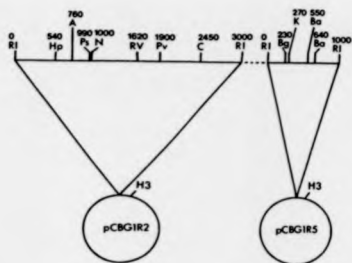
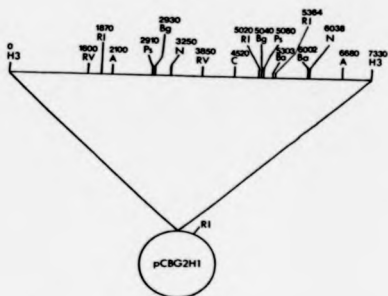


Figure 12



Figures 11-15 Restriction maps of lectin subclones

The following abbreviations have been used to denote the restriction sites present in these plasmids:- A, Acc I; Bs, Bam HI; Bg, Bgl II; C, Cla I; H1, Eco HI; HV, Eco RV; N2, Hinc II; H3, Hind III; Hp, Hpa I; K, Kpn I; M, Mco I; Pa, Pst I; Pw, Pvu II; Sa, Sal I; Sp, Sph I; X, Xba I.

HB

- 1) Some extrapolations have been made from comparisons with the lectin cDNA restriction maps. Thus those restriction site coordinates quoted to an accuracy of 1 bp have been assumed to occur at the same position as in the cDNAs.
- 1i) A comparison between the restriction maps of the genomic clones and the lectin cDNAs is shown in Figure 15.
- 1ii) The vector portions of the clones shown in Figures 11-14 are not drawn to scale, but have been included for the purpose of indicating the orientation of the inserts. Thus the orientation of one of the pUC8 polylinker sites has been shown in each case.

Figure 11 Restriction maps of pCBG1B2 and pCBG1B5

The plasmids pCBG1B2 and pCBG1B5 were obtained by separately subcloning two Bam HI fragments, of respective sizes 3.0 Kb and 1.0 Kb, from lambda clone 1 into pUC 8. The restriction maps of these two plasmids show that they contain two different portions of a putative lectin gene. The dotted line indicates the position of a putative 344 bp fragment which spans the A chain, linker and B chain regions of the lectin cDNAs but which is not present in either of the two subcloned fragments shown. Presumably this fragment links the pCBG1B2 and pCBG1B5 inserts in the R. communis genome and therefore lambda clone 1.

Figure 12 Restriction map of pCBG2H1

The plasmid pCBG2H1 was obtained by subcloning a 7.3 Kb Hind III fragment containing a putative lectin gene from lambda clone 3 into pUC 8 (Table 6).

Figure 13

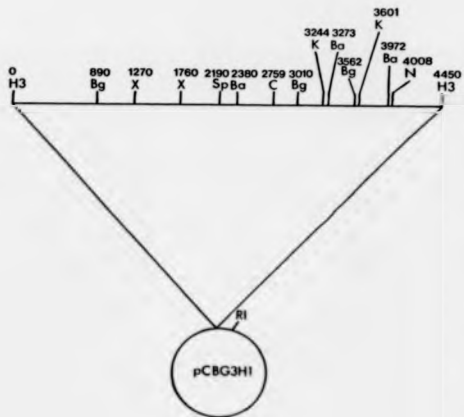


Figure 13 Restriction map of pCBG3H1

The plasmid pCBG3H1 was obtained by subcloning a 4.5 Kb Hind III fragment containing a putative lectin gene from lambda clone 10 into pUC8 (Table 6).

A comparison between the pCBG3H1 restriction map and those of the other subclones, plus the lectin cDNAs, is shown in Figure 15.

The remaining lectin-positive lambda clone chosen for further analysis was clone 7. In the Southern blotting experiment shown in Figure 9, it was demonstrated that the genomic insert in lambda clone 7 contains two hybridising Hind III fragments and three hybridising Eco RI fragments. The sizes of the latter seemed to indicate that two putative lectin genes might be present in the insert. In total, the insert in lambda clone 7 was found to contain four Hind III fragments, as judged by the pattern of bands seen on an ethidium bromide stained gel. These four fragments occur in two close doublets, each containing one of the hybridising fragments. All four fragments were separately subcloned in to pUC8 to produce the plasmids pCBG4H1, pCBG4H2, pCBG4H3 and pCBG4H4 (again the individual numbers assigned to the plasmids correspond to the order of size of their inserts). Restriction mapping confirmed that two putative lectin genes are present in the lambda clone 7 insert, since the plasmids pCBG4H2 and pCBG4H4 were both shown to contain inserts with lectin gene-like restriction maps. The pCBG4H2 insert contains a putative lectin gene plus approximately 770 bp of 5' flanking region and 200 bp of 3' flanking region. The pCBG4H4 insert contains a second putative lectin gene

which is also flanked at its 3' end by approximately 200 bp of DNA, but which is incomplete at its 5' end. The restriction maps of these two plasmids are shown in Figure 14. Presumably the two putative lectin genes are closely linked in the R. communis genome. The pCBG4H4 insert was not characterised further, but the putative lectin gene in pCBG4H2 was subsequently investigated by both DNA sequencing and RNase protection.

In order to assess the restriction site diversity of the R. communis lectin gene family, the restriction maps of all the putative lectin genes were aligned, along with the ricin and RCA I cDNA restriction maps. The combined data is shown in Figure 15 and its implications are discussed in section IV.2.

Figure 14

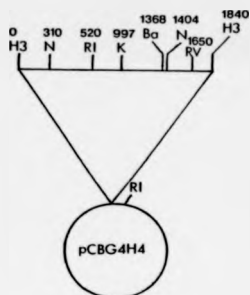
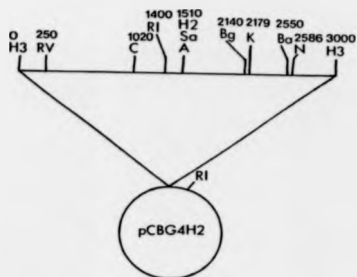


Figure 14 Restriction maps of pCBG4H2 and pCBG4H4

Two Hind III fragments, of approximate sizes 3.0 Kb and 1.8 Kb, were separately subcloned from lambda clone 7 into pUC8 to produce pCBG4H2 and pCBG4H4 respectively (Table 6). The restriction maps of the subcloned fragments show that a putative lectin gene is present in each, although the putative lectin gene in pCBG4H4 appears to be incomplete at its 5' end.

[illegible]

Figure 15 Comparison of restriction sites in the
lectin gene subclones and outline of
sequencing and RNase protection strategies

The restriction maps of the lectin genomic clones and lectin cDNA clones (Roberts et al, 1985), have been aligned so that any conserved restriction sites may be identified. Some of the genomic clones (eg pCBG2H1) are not represented in their entirety.

Beneath the restriction maps is a scale diagram showing the positions of the signal sequence (S), A chain sequence (A), linker sequence (L) and B chain sequence in the lectin cDNA clones. The extent of the 5' and 3' untranslated regions of the pCBG3H1 ricin gene is also shown.

Arrows have been used to show the location and direction of the DNA sequences determined. Regions subjected to M13 dideoxy-sequencing are denoted by arrows with continuous tails, whilst those regions subjected to Maxam/Gilbert sequencing are denoted by arrows with broken tails. Box symbols have been used to indicate the regions in the pCBG2H1, pCBG3H1 and pCBG4H2 clones which were subcloned into transcription vectors for RNase protection analysis.

SECTION III.4. SELECTIVE NUCLEOTIDE SEQUENCE ANALYSIS OF CLONES

Four of the five putative lectin genes identified in the plasmid subclones were analysed further by DNA sequencing. Only the pCBG4H4 subclone from lambda clone 7 was omitted at this stage as it did not appear to contain a full-length gene. The 5' end of this putative gene may well be present in the lambda clone 7 insert and therefore pCBG4H1 or pCBG4H3, but due to a shortage of time, no further investigation was made.

Figure 15 outlines the sequencing strategy used for each of the clones. The only putative lectin gene which was completely sequenced was that present in pCBG3H1 (lambda clone 10). Since this putative gene has an identical restriction map to that of the ricin cDNA (see Figure 15), it seemed highly likely that pCBG3H1 contains a functional ricin gene.

Although a complete nucleotide sequence of the putative lectin gene in lambda clone 1 (pCBG1R2/pCBG1R5) was not determined, nearly all of the region of interest was sequenced. The only region not sequenced was the proposed 344 bp Eco RI fragment which was not identified in the Southern blot (Figure 9) and therefore not subcloned. The

remaining sequence was determined via the pCBG1R2 and pCBG1R5 plasmid subclones. Both the pCBG1R2 and pCBG1R5 nucleotide sequences revealed the presence of frameshift mutations in the lectin "coding" region, thus demonstrating that lambda clone 1 contains a lectin pseudogene. For this reason no attempt was made to determine the sequence of the "missing" fragment.

Owing to a shortage of time, very little sequencing was carried out on the putative lectin genes in pCBG2H1 and pCBG4H2 (lambda clones 3 and 7 respectively). In both cases, a small region containing codons which specify different amino acids in the respective lectin cDNA clones was chosen. Thus it could be determined whether the putative genes contain ricin- or RCA I-like sequences, at least within the regions investigated.

All the DNA sequence data presented was obtained using M13 dideoxy-sequencing, with the exception of some of the sequence data from the lambda clone 1 (pCBG1R2/pCBG1R5) pseudogene, which was obtained using the Maxam and Gilbert sequencing method. The regions involved are shown in Figure 15. Table 7 lists the M13 clones constructed for sequencing and the corresponding genomic fragments subcloned in each case. Table 8 details the oligonucleotide primers used and their points of annealing in the ricin (pCBG3H1) sequence. For the purpose of

Table 7

Name of construct	Source of cloned DNA fragment	Restriction enzymes used to generate cloned fragment	Size of subcloned fragment	M13 vector used	Restriction enzymes used to generate suitable cohesive ends on vector
JT 1	pCBG185	<u>Eco</u> RI, <u>Bgl</u> II	226 bp	mp18	<u>Eco</u> RI, <u>Bam</u> HI
JT 2	pCBG182	<u>Eco</u> RI, <u>Pvu</u> II	1900 bp	mp18	<u>Eco</u> RI, <u>Sma</u> I
JT 3	pCBG182	<u>Eco</u> RI, <u>Pvu</u> II	1100 bp	mp18	<u>Eco</u> RI, <u>Sma</u> I
JT 4	pCBG182	<u>Eco</u> RI, <u>Pvu</u> II	1100 bp	mp9	<u>Eco</u> RI, <u>Sma</u> I
JT 5	pCBG3M1	<u>Xba</u> I, <u>Eco</u> RI	2690 bp	mp18	<u>Xba</u> I, <u>Eco</u> RI
JT 6	pCBG3M1	<u>Bam</u> HI, <u>Hind</u> III	2380 bp	mp19	<u>Bam</u> HI, <u>Hind</u> III
JT 7	pCBG3M1	<u>Bgl</u> II, <u>Sph</u> I	1300 bp	mp18	<u>Bam</u> HI, <u>Sph</u> I
JT 8	pCBG3M1	<u>Sph</u> I, <u>Bgl</u> II	820 bp	mp18	<u>Bam</u> HI, <u>Sph</u> I
JT 9	pCBG3M1	<u>Xba</u> I, <u>Hind</u> III	2690 bp	mp18	<u>Hind</u> III, <u>Xba</u> I
JT 10	pCBG2M1	<u>Eco</u> RV, <u>Eco</u> RI	1170 bp	mp19	<u>Sma</u> I, <u>Eco</u> RI
JT 11	pCBG4M2	<u>Hind</u> III, <u>Bgl</u> II	2140 bp	mp19	<u>Hind</u> III, <u>Bam</u> HI

Table 7 M13 subclones constructed for sequencing

The table opposite details the cloning strategy used to prepare the M13 sequencing constructs used in this study. The restriction maps of the plasmids from which the subcloned DNAs were obtained are shown in Figures 11 to 14, whilst the regions sequenced are identified in Figure 15.

Table 2

Name	Nucleotide sequence	Point of annealing in ricin sequence	Source of oligonucleotide (coding or non- coding)	W13 templates sequenced
Universal	GTAAACCAACCGCCACT	-	Not applicable	JT 1, JT 2, JT 3, JT 4, JT 7, JT 8, JT 9, JT 10, JT 11
LE4	TATAATTCGCTATTG	118*	Non-coding	JT 4
CL301	TTTGCTTATAACCAACCG	259*	Coding	JT 5
CL303	TCTTTGCTTATAAGGCC	907*	Coding	JT 5
CL304	CGCTGGCAAAATATGG	1207*	Coding	JT 5
CL305	AATTGCTTACAAGTGA	1507	Coding	JT 5
CL306	CAGATTACTCTCTTGC	5**	Coding	JT 5
CL581	ATGAAACCGGAGGAAA	1*	Coding	JT 5
CL582	CCCAACCAATCCCAAT	112*	Coding	JT 5
CL583	ACCGTCTGCAATAGC	336*	Coding	JT 5
CL584	TGCTAATCTGAGAGAAA	495*	Coding	JT 5
CL585	AGATCCTAGCGTAATTA	705*	Coding	JT 5
CL586	GCAAGACGTAATGCTT	791*	Coding	JT 5
CL587	TGATGTTAGGGATCGAA	1005*	Coding	JT 5
CL588	GATCTAATGGAAAGTGT	1115*	Coding	JT 5
CL589	TTATGCGTTAGTCAAG	1114*	Coding	JT 5
CL590	ATAGTGGACAAGTATGG	1406*	Coding	JT 5
CL591	CGATGATGTTCAACAA	1585*	Coding	JT 5
CL592	ATTGCTCTGTGCTATTCT	134**	Coding	JT 5
CL593	TTTCTCCCGGTTTCAT	1*	Non-coding	JT 6

* refer to Figure 16.b

** refer to Figure 16.c

Table 8 Oligonucleotides used for M13 dideoxy-
sequencing

The oligonucleotide primers used in this study for sequencing the lectin genomic clones are listed opposite.

sequencing the pCBG3H1 (lambda clone 10) ricin gene, a variety of oligonucleotides homologous to the ricin cDNA were used as primers. The sequences obtained using these primers were overlapping, making it possible for the coding region of the gene to be completely sequenced using a single M13 clone.

Figure 16 shows the nucleotide sequence determined for the pCBG3H1 (lambda clone 10) ricin genomic clone. The sequence is compared with the two other published ricin sequences, namely that of a ricin cDNA (Lamb et al., 1985) and a ricin genomic clone (Halling et al., 1985).

Figure 17 shows the partial nucleotide sequence determined for the lambda clone 1 (pCBG1R2/pCBG1R5) lectin pseudogene. In order to establish whether this pseudogene is ricin- or RCA 1-like, the "coding" sequence is compared with those of the ricin and RCA 1 cDNAs (Roberts et al., 1985).

Figure 18 shows the nucleotide sequence of an internal region from the putative lectin gene in pCBG2H1 (lambda clone 3), whilst Figure 19 shows the nucleotide sequence of a different internal region from the putative lectin gene in pCBG4H2 (lambda clone 7). Again, in both cases, a comparison is made with the ricin and RCA I cDNA sequences (Roberts et al., 1985).

Figure 16.a

1
pCBG3H1 CCTAAAAAAGTAAATTACTCTAATCGACATTATGAATTTTAACT-AA
pAKG A C

50
pCBG3H1 TTCCGTTTCTAATTATTAATTATTTTCGTTAAACCAATCAATTCCCTTTAA
pAKG

100 VII VI
pCBG3H1 ACAC TGCT TATGCATA TCTGTCTCAATTTATATATGG CAT-GCAT-CIT
pAKG T T

148 IV III
pCBG3H1 CCTTATTAATTATTAAGTT-A-TTTT TATGATCAAGT ATTGTGGTTTT
pAKG C C

196 II V I
pCBG3H1 CTT TATATA AAAAAATGTATTAGTGTTTTCT GTATTAATT TATAAGTT
pAKG

246 ▽
pCBG3H1 CATCTTTATGAGAATGCTAATGTATTGGACAGCCAATAAAATCCAGAA
pAKG

296
pCBG3H1 TTGCTGCAATCAAAG
pAKG G

Figure 16.a Nucleotide sequence of 5' flanking region
and 5' untranslated region of the pCBG3H1
(lambda clone 10) ricin gene

The nucleotide sequence shown covers a 310 bp region immediately upstream from the coding region of the pCBG3H1 (lambda clone 10) ricin gene (see Figure 15). The pCBG3H1 sequence is compared with the sequence of a ricin genomic clone, pAKG, isolated by Halling *et al.* (1985). The nucleotide sequence of the pAKG clone is shown only where the two sequences differ. Dashes represent nucleotides which are present in one of the clones only. An arrowhead symbol indicates the most likely cap site position identified by RNase protection (Figure 25). Six regions of interest are highlighted as boxes. Boxes I and II indicate the respective positions of two TATA sequences, whilst box III indicates the position of a potential CAAT/AGGA box analogue. Box V shows the position of an 18 bp sequence which is duplicated almost perfectly upstream in box IV. Finally, box VI shows the position of a 14 bp region which is highly homologous to a sequence in the promoter region of a castor bean 2S albumin gene and includes a CATGCATC sequence which closely resembles the RY repeat found in many seed-specific promoters (Dickinson *et al.*, 1988). Another sequence resembling an RY repeat is found further upstream in box VII.

Figure 15.b



Figure 16.b Nucleotide sequence of the coding region
of the pCBG3H1 (lambda clone 10) ricin
gene and deduced primary sequence of the
encoded polypeptide

The nucleotide sequence shown covers the coding region of the pCBG3H1 ricin gene subcloned from lambda clone 10 and is a continuation of the pCBG3H1 sequence in Figure 16.a (see Figure 15). The deduced primary sequence of the encoded polypeptide is shown above the nucleotide sequence. A comparison is made between the nucleotide sequence of the pCBG3H1 ricin gene and the sequences of the pAKG ricin genomic (Halling *et al*, 1985) and pRCL617 ricin cDNA (Lamb *et al*, 1985) clones respectively. Mismatches only are shown in the pRCL617 and pAKG sequences. Nucleotide coordinates are shown to the left of each line, whilst primary sequence coordinates are shown to the right.

Figure 16.c

1
pCBG3H1 TAGACAGATTACTCTCTTGCACTGTGTATGTCTGCTATGAAAATAGATG
pAKG C

51
pCBG3H1 GCTTAAATAAAAGGACATTGTAAATTTGTAACTGAAAGGACAGCAAGT
pAKG

101
pCBG3H1 TATTGCAGTCCAGTATCTAATAAGAGCACAACTATTGTCTTGTGCATTCT
pAKG

151 ▽ ▽
pCBG3H1 AAATTTATGGATGAATTGTATGAATTAAAGCTAATTATTTGGTCATCAGA
pAKG A

201
pCBG3H1 CTTGATATCTTTTGAATAAAATAAATAA-TGTTTTTCAAACCTTATA
pAKG A

250
pCBG3H1 AATCTA-TGAATGATATGAATATAA-TCCGGAGACTAGTCAATCTTTTAT
pAKG A A A

298
pCBG3H1 GTAATTCTATGATGATAAAAGCTT
pAKG

Figure 16.c Nucleotide sequence of the 3'
untranslated region and 3' flanking
region of the pCBG3H1 (lambda clone
10) ricin gene

The nucleotide sequence shown is a continuation of the pCBG3H1 sequence in Figure 17.b (see Figure 15). The pCBG3H1 sequence is compared with the corresponding sequence of the pAKG ricin genomic clone (Halling et al. 1985). Mismatches and deleted nucleotides are represented as in Figure 17.a.

The positions of two potential poly(A) sites identified by RNase protection (Figure 26) are indicated by arrowheads, whilst the boxed regions show the positions of two potential poly(A) signals.

AGTTTAACCTACTTAATGGAGCATTATTTATCAGGGAAATAAAATCATT

51
AGTTAAGGGACTAGATCCTATTATTAATCCCTATCCCGAATTTTAACGGC

101
AAGAAGAACTAAAAAAAATAAGTTAGTTTCATATGATAATAATAATAAT

151
ATAATAATAATAATAATAATAATAATAATAATAATAATCTATTTTTT

201
ATAAGTATTCATTCTAATTTACGAATATTTTCATTGAACCAATAATTTTT

251
TTAAAAAATATTGTTTATGATATATTATTTTCAGTTTATTTTCGTGATA

301
CATCTTCTCTATTAATTTTATAAGTTCATTCTTTTATCAAAGTATTTTT

351
ATTTTCTTTTAAAGAAATGTTTATATATATCTTATGGATCATCTATG

401
TTTATGAGAATATTAATGCATTTGGACAGCTAGAATTGTTGCAATCAAAG
(pCBG3H1) AGAATTGCTGCAATCAAG

1
AGTTTAACCTACTTAATGGAGCATTATTTATCAGGGAAAATAAAATCATT

51
AGTTAAGGGACTAGATCCTATTATTAATCCCTATCCCCAATTTTAACGGC

101
AGAAGAAGCTAAAAAATAAAGTTAGTTTCATATGATAATAATAATAAT

151
AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATACTATTTTTTTT

201
ATAAGTATTCATTTCTAATTTACGAATATTTCAATGAACCAATAATTTT

251
TAAAAAATATTGTTTATGATATATTATTTCTCAGTTTATTTTCGTCATA

301
CATCTTCTCTATTAATTTTATAAGTTCATTCTTTTATCAAAGTATTTTT

351
ATTTTCTTTTAAAGAAATGTTTATATATATATTCTTATGGATCATCTATG

401
TTTATGAGAATATTAATGCATTTGGACAGCTAGAATTGTTGCAATCAAAG
(pCBG3H1) AGAATTGCTGCAATCAAGG

Figure 17.a Partial nucleotide sequence of a lectin
pseudogene from lambda clone 1: Sequence
of 5' flanking region and region sharing
homology with the ricin mRNA 5' leader
sequence

The nucleotide sequence opposite covers a 450 bp region in pCBG1R2 which precedes the lambda clone 1 lectin pseudogene (see Figure 15).

A short region of homology with the pCBG3H1 ricin gene 5' untranslated region occurs as shown. A repetitive sequence consisting of the trinucleotide ATA repeated 18 times is highlighted within a box.

Figure 17.b

pCB182
pBCL617
pBCL521

1
ATGCAACTGGGAGGAAATACTATTATAATTGGATATATTCACTGGCAACATGGCTTTGG
A C G A C G G

61
TTTGGATCTACCTCAGGGTCTCTATTCACTTTAGCGGATAACACATATTACCAATATAT
C T A A A
A
A
A

pCB182
pBCL617
pBCL521

121
AAAGCTCATCTACTGGGGTGGCACTGGCCAAAAGCTACACAACTTTATCGAT**GCTG
XX

pCB182
pBCL617
pBCL521

181
TCTCCAGCATCTTCACTGCACTGGCAGATGATGACACAAGAAATACCAGTGTTCGCAAAACA
XX
XX

pCB182
pBCL617
pBCL521

241
GAGTTCATTTCGCTATAAGCCACGATTTCTTTGATTCAACTCTCAAACTCAGGACAGCC
XX
XX

pCB182
pBCL617
pBCL521

301
ATTCTCTACATTCGGCGATGGATCTGCCAAATGCAATCTGGTGGGTTACCGCTCTCGAAA
XX
C G C

pCB182
pBCL617
pBCL521

361
TTTGCAATATTTCTTCTCTCAAACTTCTGAAGATGCAGAGACCAATTAATCTCTCTTTT
AG C A CAG
AG C A C A A

pCB182
pBCL617
pBCL521

421
CATGAGCTCTCAAAATCCATATACATTCACCTTTGGTGGAAACTATGATAGACTTCAACA
CT G T T T T T
G T T T T T

pCB182
pBCL617
pBCL521

481
ACTT---CGAGGTTTCAGACAAAAATACGAGCTGGCAAAATGATCCATCGGAGGATGCTAT
CCT TAA C T T G C G G T A C

pCB182
pBCL617
pBCL521

538
CTCAGCGCTTATATTATAGTACTGGTGGCACTACGCTTCCAATCTACGCTGCTTCATT
C T A G C C C

pCB182
pBCL617
pBCL521

598
CATTTGTTTCATCCAAATGGCTTCAGAAGCACTAAGATTCCAATATATCGGAGGGAAT
T AA AT C T-
T G C C T-

pCB182
pBCL617
pBCL521

658
ACGCACAAGAATTC
C C A A
G C G A

Figure 17.b Partial nucleotide sequence of a lectin
pseudogene from lambda clone 1: Sequence
of region sharing homology with ricin/
RCA I signal peptide and A chain coding
sequences

The 672 bp nucleotide sequence shown is a continuation of the pCBG1R2 sequence shown in Figure 17.a. This region shares homology with the ricin and RCA I signal peptide and A chain coding sequences (see Figure 15). A comparison is made between the pCBG1R2 sequence and those of the pRCL617 (ricin) and pRCL521 (RCA I) cDNA clones (Roberts et al, 1985). The pRCL617 and pRCL521 sequences are indicated only at positions where they differ from the pCBG1R2 sequence.

A "scrambled" region containing sequence which is unrecognisable as lectin coding sequence is indicated by the letter X beneath the relevant pCBG1R2 nucleotides. Three nucleotide positions within the "scrambled" region are marked by asterisks, since the residues in question were not identified. The arrowhead marked "FS" shows the position of a frameshift mutation in the pseudogene.

Figure 17.c Partial nucleotide sequence of a lectin
pseudogene from lambda clone 1: Sequence
of a region sharing homology with the
ricin/RCA I B chain coding sequence

The nucleotide sequence shown spans 710 bp at the 3' end of the B chain region of the lambda clone 1 pseudogene. This region is contained in the pCBG1R5 subclone (see Figure 15). As with Figure 17.b, a comparison is made between the pseudogene sequence and those of the pRCL617 (ricin) and (RCA I) cDNA clones (Roberts et al., 1985). A frameshift mutation is indicated by an arrowhead marked "FS".

Figure 17.d

1
TGTATTGATTATCCTCCTAGTCTCTTGCACTTGTAGTGTGTTGTATGTCCT

51
GTTATGGTGCATAGTTGCTTAATAAAAGGACATTGTAAATTAAACCTG

101
TAACTGAAAGGACAACAAGTTATTGCAGTCCAGTATTGAATAAGAGTAAA

151
AGCACAGGCTATATAGTCTTGTGCTTTCTTTACTTATGACAAAGATGAA

201
TTGTATGATTATGTTAATTATCTCTTAGGTGTGCTTTTAAGTGACATC

251
CACTTACATTIATTTATACATATATACCATTGTAAGCCAACCTTAGCATTG

301
AAAACCTCCGCAACAATACTTTATAATCATGCAACAATCATGAATCTTGAA

351
TTC

Figure 17.d Partial nucleotide sequence of a lectin
pseudogene from lambda clone 1: Sequences
downstream from the region sharing
homology with the ricin/RCA I coding
sequence

The nucleotide sequence shown is a continuation of the pCBG1R5 sequence in Figure 17.c. The region in question occurs immediately 3' to the B chain region of the lambda clone 1 lectin pseudogene and extends as far as the Eco RI site bordering the pCBG1R5 insert (see Figure 15). Two elements matching the respective sequences of the two putative poly (A) signals in the lectin cDNA 3' untranslated regions (Roberts et al., 1985) are highlighted as boxes.

Figure 18

1
 pCBG2H1 GGAGGCCTGAGAGAAAATATCGACCTGGCAAATGATCCACTGGAG
 pRCL617 TAA T T G G A
 pRCL521 T T T G C G T A

46
 pCBG2H1 GATGCTATCTCAGCGCTTTATTATTATAGTACTGGTGGCACTCAG
 pRCL617 G C C T
 pRCL521 C T

FS
 ▽
 pCBG2H1 GCTTCCAACCTCTAGCTCGTTCATTCATTGTTTGCAATCCAAATGGC
 pRCL617 - G C T AA AT
 pRCL521 -A G C T G AT

FS
 136
 pCBG2H1 TTCAGAAGCACTAAGATTCCAATATATCGTAGGGA
 pRCL617 GC T -
 pRCL521 GC G C T -

Figure 18 Nucleotide sequence of an internal region
from the pCBG2H1 (lambda clone 3) lectin
pseudogene

The sequence shown covers a 170 bp region on the 5' side of the Eco RI site in the A chain region of the pCBG2H1 (lambda clone 3) lectin pseudogene (see Figure 15). A comparison is made between the pCBG2H1 sequence and those of the pRCL617 (ricin) and pRCL521 (RCA I) cDNA clones (Roberts et al., 1985). The pRCL617 and pRCL521 sequences are indicated only at positions where they diverge from the pCBG2H1 sequence. The arrowheads marked "FS" show the positions of two frameshift mutations in the pseudogene.

Figure 19

	1	TATGTGATGATCTATGATTGGAATACTGCTGGAAC	TGATGCGACC
pCBG4H2			
pRCL617		C	C
pRCL521		C G G A A C G C AC GT G C	

	46	CGCTGGCAAATATGGGATAATGGAACCATATAAATCCCAGATCT
pCBG4H2		
pRCL617		
pRCL521		T C C C

Figure 19 Nucleotide sequence of an internal region
from the pCBG4H2 (lambda clone 7) lectin
pseudogene

The sequence shown covers a 90 bp region on the 5' side of the Bgl II site in the B chain region of the pCBG4H2 (lambda clone 7) lectin pseudogene (see Figure 15). The pCBG4H2 sequence is compared with the nucleotide sequences of the pRCL617 (ricin) and pRCL521 (RCA I) cDNA clones (Roberts et al, 1985). The pRCL617 and pRCL521 sequences are indicated only at positions where they diverge from the pCBG4H2 sequence.

SECTION III.5. LEVELS AND PATTERNS OF LECTIN GENE
TRANSCRIPTION

III.5.A. Isolation of total RNA

In order to characterise the developmental and tissue-specific nature of lectin gene transcription in R. communis, total RNA was prepared from leaves, roots and seeds of different developmental stages. The developing seeds used were categorised into six arbitrary groups, A to F, as was previously done by Roberts and Lord (1981B.). Total RNA was also extracted from dry seeds and the endosperm of germinating seeds. No attempt was made to excise the cotyledons from the developing seeds, although the testas of the seeds from stages D to F were removed. The levels of lectins in castor bean cotyledons, which contribute only a small part of the dry seed weight, are less than 1% of those observed in the endosperm (Harley and Beevers, 1986). Therefore, the levels of lectin gene transcripts in the RNA samples obtained will be essentially identical to those in the endosperm tissue itself. Figure 20 shows the appearance of the seeds in each developmental stage. For the purpose of isolating RNA from seed tissues, the method of Keller and Taylor (1976), as described by Roberts and Lord (1981B.), was used. Attempts at isolating total RNA from the leaf and

Figure 20



Figure 20 Stages of *R. communis* seed development

R. communis plants were grown to maturity as described in section I.2.A. Fruits were harvested at different stages of development and dissected. Seeds were categorised into six developmental stages (A to F) on the basis of their size and colouration. Stages A to C precede the period of testa formation which starts at stage D and is complete by stage F. Stage DS denotes the dry seed.

root tissues using this method were, however, unsuccessful. Instead, the method of Logemann et al (1986) was used. Although the RNA yields obtained were low and slight degradation may have occurred, the samples obtained were considered to be of sufficient quality for use in Northern and RNase protection analysis.

III.5.B. Northern analysis of lectin gene transcription

Northern analysis of transcription was carried out as described in Figure 21. A Northern blot was probed with a ricin cDNA and washed at low stringency, so that both ricin and RCAI gene transcripts would be detected. The resulting autoradiograph is shown in Figure 21. As is discussed in section IV.3., the autoradiograph shows that the mRNAs of the castor bean lectin genes accumulate during the latter stages of seed development.

III.5.C. RNase protection analysis of transcription using probes from the pCBG3H1 (lambda clone 10) ricin gene

As a means of monitoring gene transcription, nuclease protection offers three main advantages over Northern blotting. Firstly, the sensitivity of mRNA detection obtained by nuclease protection is greater than that obtained by Northern blotting. Secondly, nuclease

Figure 21

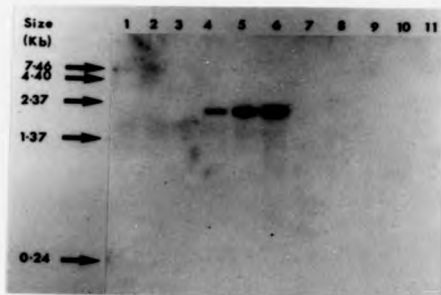


Figure 21 Northern analysis of lectin transcript
levels in different tissues and stages of
seed development

Total RNA was isolated from castor bean root and leaf tissue and from seeds of various developmental stages (see Figure 20). A 10 ug aliquot was removed from each preparation and the samples were electrophoresed on a 50% formamide, 1.2% (w/v) agarose, 0.1 x TEP gel alongside RNA size markers. The gel was Northern blotted onto Hybond-N and the membrane was probed with a radiolabelled ricin cDNA fragment prepared by the oligolabelling. Following hybridisation, the membrane was washed at low stringency (section II.4.D.) and autoradiographed overnight. The position of the RNA size markers is indicated by the arrows. The sources of tissue used to obtain the castor bean RNA samples are as follows:- track 1, seed developmental stage A; track 2, seed developmental stage B; track 3, seed developmental stage C; track 4, seed developmental stage D; track 5, seed developmental stage E; track 6, seed developmental stage F; track 7, dry seeds; track 8, endosperm tissue harvested from germinating seeds 3 days after imbibition; track 9, endosperm tissue harvested from germinating seeds 7 days after imbibition; track 10, mature leaf tissue; track 11, root tissue.

protection analysis is sequence-specific, so the transcriptional activity of a specific gene of interest may be monitored. Sometimes the transcriptional activity of several different genes may be determined in a single reaction. Thirdly, nuclease protection offers a means of identifying the position of the 5' and 3' termini of the mRNA under investigation. The S1 nuclease mapping technique (Weaver and Weissmann, 1979) is generally considered to resolve mismatches of 2 bp in the DNA probe: mRNA hybrids formed. The sensitivity obtained by RNase protection (Kreig and Melton, 1986) is, however, considered sufficient to resolve single mismatches. In addition, the experimental procedure used for RNase protection is slightly simpler. Thus this method was chosen for use in this study.

The RNase protection experiment carried out using probes obtained from the pCBG3H1 (lambda clone 10) ricin gene enabled identification of potential cap and poly (A) sites as well as a more sensitive analysis of the developmental and tissue specific transcriptional pattern of the gene. Two different RNA probes, one spanning the 5' end of the gene and the other spanning the 3' end of the gene were used. The probes were obtained by in vitro transcription from the recombinant plasmids pGEM-blue3JT5 and pGEM-blue4JT2 respectively. Table 9 details the genomic fragments subcloned to produce these constructs, whilst

Table 3

Name of construct	Source of cloned DNA fragment	Restriction enzymes used to generate cloned fragment	Size of subcloned fragment	Transcription vector used	Restriction enzymes used to generate suitable cohesive ends on vector
pGEN-blue3JT1	pCBG3H1	<u>Sph</u> I, <u>Cla</u> I	618 bp	pGEN-blue3	<u>Acc</u> I, <u>Sph</u> I
pGEN-blue4JT2	pCBG3H1	<u>Bam</u> HI, <u>Hind</u> III	393 bp	pGEN-blue4	<u>Hind</u> III, <u>Bam</u> HI
pGEN-blue3JT3	pCBG2H1	<u>Bgl</u> II, <u>Eco</u> RI	324 bp	pGEN-blue3	<u>Eco</u> RI, <u>Bam</u> HI
pGEN-blue3JT4	pCBG4H2	<u>Cla</u> I, <u>Eco</u> RI	380 bp	pGEN-blue3	<u>Eco</u> RI, <u>Acc</u> I
pGEN-blue3JT5	pCBG3H1	<u>Sph</u> I, <u>Bam</u> HI	239 bp	pGEN-blue3	<u>Bam</u> HI, <u>Sph</u> I

Table 9 Clones constructed for RNase protection
analysis

DNA fragments were subcloned into transcription vectors from three of the putative lectin genes characterised in this study. The cloning strategy used and the nomenclature of the subclones is described opposite.

the positions of the regions in question are shown in Figure 15. The total RNA samples used in the RNase protection experiments were the same as those used for Northern blotting. The result obtained when the probe spanning the 5' end of the pCBG3H1 (lambda clone 10) ricin gene was used is shown in Figure 22, whilst the corresponding result obtained using the 3' terminal probe is shown in Figure 23.

III.5.D. RNase protection analysis of transcription using internal probes from the putative lectin genes in pCBG2H1 (lambda clone 3) and pCBG4H2 (lambda clone 7)

The RNase protection experiments carried out using probes from the putative lectin genes in pCBG2H1 (lambda clone 3) and pCBG4H2 (lambda clone 7) were performed in parallel with the DNA sequencing described in Figures 18 and 19. Since it was not known at the time whether the two putative lectin genes were functional, it was decided that an internal fragment would be used as a probe in each case. The identification of RNase-protected fragments sharing 100% homology with the probes is straightforward with this approach, since the expected size of the fragments is the same as that of the respective DNA inserts in the transcription plasmid constructs.

Figure 22

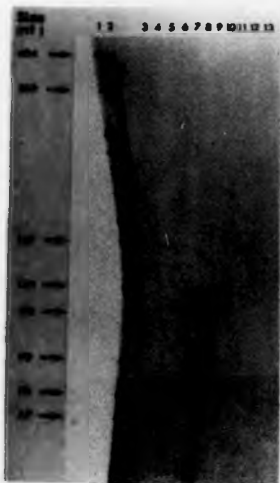


Figure 22 RNAse protection analysis of transcripts
using a probe spanning the 5' end of the
pCBG3H1 (lambda clone 10) ricin gene

10 ug aliquots from each of the total RNA preparations used in Northern blotting (Figure 21) were analysed by RNAse protection for transcripts homologous to the pCBG3H1 (lambda clone 10) ricin gene. The probe used was produced by in vitro transcription from the plasmid pGEN-blue3JT5 (see Table 9), which had been linearised with Hind III. The 237 bp cloned insert in this plasmid extends from the Sph I site in the ricin gene 5' flanking region to the Bam HI site in the signal sequence (see Figure 15). A sample of Hpa II-digested, ³²P-radiolabelled pBR322 DNA markers is shown in track 1. Track 2 contains an undigested sample of the purified RNA probe. Tracks 3 to 13 show the products of the RNAse protection assays. The source of castor bean RNA for each sample is as follows:- track 3, seed tissue of developmental stage A; track 4, seed tissue of stage B; track 5, seed tissue of stage C; track 6, seed tissue of stage D; track 7, seed tissue of stage E; track 8, seed tissue of stage F; track 9, dry seed tissue; tracks 10 and 11, endosperm tissue harvested from germinating seeds 3 and 7 days respectively after imbibition; track 12, mature leaf tissue; track 13, root tissue.

Figure 23



Figure 23 RNAse protection analysis of transcripts using a probe spanning the 3' end of the pCBG3H1 (lambda clone 10) ricin gene

10 ug aliquots from each of the total RNA preparations used in Northern blotting (Figure 21) were analysed by RNAse protection, again for transcripts homologous to the pCBG3H1 (lambda clone 10) ricin gene. The probe used in this case was produced by in vitro transcription from the plasmid pGEM-blue4JT2 (see Table 9), which had previously been linearised with Eco RI. The 391 bp cloned insert in this plasmid extends from the Bam HI site in the ricin B chain region to the Hind III site in 3' flanking region (see Figure 15). Track 1 contains a sample of Hpa II-digested, ³²P-radiolabelled pBR322 DNA markers. Track 2 contains an undigested sample of the purified RNA probe. Tracks 3 to 13 show the products of the RNAse protection assays. The source of castor bean RNA for each sample is as follows: track 3, stage A seed tissue; track 4, stage B seed tissue; track 5, stage C seed tissue; track 6, stage D seed tissue; track 7, stage E seed tissue; track 8, stage F seed tissue; track 9, dry seed tissue; tracks 10 and 11, endosperm tissue harvested from germinating seeds 3 and 7 days respectively after imbibition; track 12, mature leaf tissue; track 13, root tissue.

The probe derived from the putative lectin gene in pCBG2H1 (lambda clone 3) was obtained by in vitro transcription from the recombinant plasmid pGEM-blue3JT3, whilst the corresponding probe from the putative lectin gene in pCBG4H2 (lambda clone 7) was obtained in the same way from the plasmid pGEM-blue3JT4. Details of the genomic fragments subcloned to produce these constructs are given in Table 9 whilst the positions of the regions in question are indicated in Figure 15. In all other respects, the RNase protection experiments were performed in the same way as those described in the previous section (section III.5.B).

Figure 24 shows the result of the RNase protection experiment carried out using the probe derived from the putative lectin gene in pCBG2H1 (lambda clone 3). It can be readily seen that none of the RNase-protected fragments evident on the autoradiograph are of a size approaching that of the pGEM-blue3JT3 insert. Thus the lectin gene-like element in pCBG2H1 (lambda clone 3) is presumably a non-functional pseudogene. The sequence data shown in Figure 18 confirms this conclusion.

In the corresponding RNase protection experiment carried out using the probe derived from the putative lectin gene in pCBG4H2 (lambda clone 7), no RNase protected species of any size could be identified. Thus the lectin gene-like

Figure 24

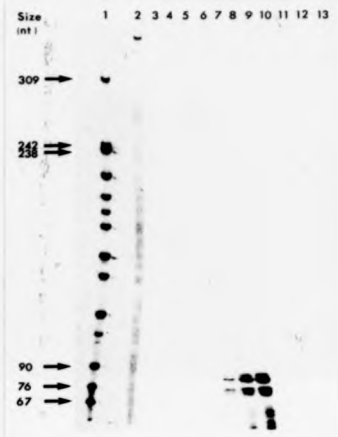


Figure 24 RNAse protection analysis of transcripts
using an internal probe from the putative
lectin gene in pCBG2H1 (lambda clone 3)

10 ug aliquots from each of the total RNA preparations used in Northern blotting (Figure 21) were analysed by RNAse protection for transcripts homologous to the potential lectin gene in pCBG2H1 (lambda clone 3). The probe used was produced by in vitro transcription from the plasmid pGEM-blue3JT3 (Table 9), which had previously been linearised with Hind III. The 324 bp cloned insert in this plasmid extends from the Bgl II site in the A chain region to the Eco RI site in the B chain region (see Figure 15). Track 1 contains a sample of Hpa II-digested, ³²P - radiolabelled pBR322 DNA markers. Track 2 contains an undigested sample of the purified RNA probe. Tracks 3 to 13 show the products of the RNAse protection assays. The source of castor bean RNA for each sample is as follows:- track 3, stage A seed tissue; track 4, stage B seed tissue; track 5, stage C seed tissue; track 6, stage D seed tissue; track 7, stage E seed tissue; track 8, stage F seed tissue; track 9, dry seed tissue; tracks 10 and 11, endosperm tissue harvested from germinating seeds 3 and 7 days respectively after planting; track 12, mature leaf tissue; track 13, root tissue.

element in pCBG4H2 must also be a non-functional pseudogene. The autoradiograph in question is not shown.

SECTION IV

DISCUSSION

SECTION IV.1. DIVERSITY OF LECTIN CLONES

IV.1.A. Restriction site diversity

A total of 17 lectin-positive lambda clones were isolated from the two R. communis genomic libraries constructed, one from the partial Eco RI library and the remaining 16 from the partial Sau 3A library.

DNA minipreparations made from each of the 17 lambda clones were analysed by Southern blotting as shown in Figures 9 and 10. The patterns of hybridising bands seen with the DNA digests from clones 1 to 9 are shown in Figure 9, whilst Figure 10 shows the corresponding hybridisation patterns of the DNA digests from clones 10 to 17. It is immediately apparent that the lectin-positive lambda clones produce hybridisation signals of appreciably different intensities, even when the filters are washed at this relatively low stringency. For this reason, the track containing the Eco RI digest from lambda clone 3 needed to be overexposed during printing in order for the bands to be visible. At the stringency used, the ricin cDNA control (C1) produces a much stronger signal than the RCA I cDNA control (C2). Hybridisation signal strength was therefore used as a criterion by which to categorise the lambda clones. It was assumed that the

lambda clones producing strong signals contain ricin-like genes. Those clones which produced weak signals were tentatively assumed to contain RCA I-like genes. This assumption was never tested, however, by reprobing the filters with an RCA I cDNA fragment.

The lambda clones were also categorised on the basis of their restriction patterns. A number of different groups of lectin clones could be distinguished in this way. Although neither of the lectin cDNAs contains a Hind III site, the RCA I cDNA differs from that of ricin in that it contains an Eco RI site. Eco RI sites appear to be present only in some of the weakly hybridising clones, suggesting that they may contain RCA I-like genes. The extra bands seen in the Eco RI digests of the strongly hybridising clones are presumed to be due to Eco RI* activity, since they did not correspond to visible bands on the ethidium bromide-stained gels and were of much lower intensity than the principal bands. This conclusion was later confirmed by restriction analysis of the subcloned DNAs.

Using the criterion described above, the lectin-positive lambda clones were classified into five groups, A to E, as described in Table 6. It should be noted that some of the clones classified in group C, which have been marked with an asterisk, differed from the rest of the group in

that they gave weaker hybridisation signals. This was apparently caused by poor DNA yields in the lambda minipreparations. This assumption was supported when it was later shown that clone 6 contains a nearly full-length ricin-like gene with an identical restriction pattern to that in clone 10 (data not shown). Since the criteria used to classify the clones were relatively simple, it is possible that some lectin genes might not be distinguished in this way. Thus the possibility that one or more of the groups contains clones of more than one lectin gene must be considered. Although this may explain the large number of clones in group C, it seems unlikely to be the case with the other groups, which each consist of only 1 to 3 clones. One clone was chosen from each group, except group E, for further analysis, as described in Table 6. Table 6 also details the plasmid subclones obtained from each lambda clone. Figures 11 to 14 show the restriction maps of the subcloned DNAs, whilst the overall restriction site diversity of the putative lectin genes and their flanking sequences is summarised in Figure 15. The fact that the genomic clone restriction maps align well with the cDNA restriction maps suggests that the castor bean lectin genes lack intervening sequences. This has already been demonstrated with the ricin genomic clone isolated by Halling *et al* (1985). The restriction maps of the inserts in two different subclones from lambda clone 7 (namely pCBG4H2 and pCBG4H4) have been shown in Figure 15,

since both were found to contain putative lectin genes with different restriction maps. The putative lectin gene in pCBG4H4 is incomplete at its 5' end, however.

On the basis of the restriction site diversity seen in Figure 15, it is apparent that a minimum of 5 putative lectin genes have been cloned. Another putative lectin gene may be present in the lambda clone 14 insert, but this remains to be determined. It is possible that lambda clone 14 might contain a genomic fragment which overlaps that in lambda clone 7, but which does not contain the pCBG4H2 and pCBG4H4 inserts in their entirety. Even if this hypothesis is correct, it appears that the complement of genes cloned is not sufficient to account for all the bands seen in the genomic blot in Figure 5. Since the R. communis lectin gene family is estimated to contain approximately 8 members, it is probable that 2 to 3 genes have not been isolated. Although the pCBG3H1 (lambda clone 10) ricin gene restriction map exactly matches that of the ricin cDNA (Lamb et al., 1985), none of the clones described has a restriction map which matches that of the RCA I cDNA (Roberts et al., 1985). Thus at least one of the genes not isolated must presumably encode an RCA I polypeptide.

Overall, the putative lectin genes show a fairly high degree of restriction site diversity. The only

restriction site common to all the clones is the Bam HI site near the 3' terminus of the B chain coding sequence in the lectin cDNAs. The Kpn I and Nco I sites in the B chain region are also highly conserved, being absent only in the pCBG2H1 and pCBG1R5 clones respectively. Overall, the restriction maps of the subclones deviate considerably from those of the cDNAs, with the exception of pCBG3H1 (lambda clone 10). As might be expected, even fewer conserved sites are present in the flanking regions. One notable exception appears to be the Hind III site which occurs just downstream from three of the five putative lectin genes, namely those in pCBG3H1 (lambda clone 10), pCBG4H2 (lambda clone 7) and pCBG4H4 (lambda clone 7). Some restriction site similarities appear to occur in the 5' flanking regions of the putative lectin genes, but the situation is somewhat more complicated. The most interesting example is in the 5' flanking regions of the putative lectin genes in pCBG1R2 (lambda clone 1) and pCBG2H1 (lambda clone 3). In both cases, a Nco I site is seen to occur approximately 600 bp upstream from an Eco RV site. These sites occur approximately 200 bp further upstream in the pCBG1R2 clone than in the pCBG2H1 clone, however. The most likely explanation is that a DNA insertion event may have taken place in the region contained in pCBG1R2 since the two genes diverged. It appears that the Eco RV site mentioned is also conserved in the pCBG4H2 clone. The only other 5' flanking region

restriction site which appears to be conserved amongst the clones is the Bgl II site seen in the pCBG2H1 and pCBG3H1 clones.

To summarise, the five putative lectin genes analysed show considerable restriction site diversity. The pCBG3H1 (lambda clone 10) restriction map matches that of the ricin cDNA, suggesting that it contains a functional ricin gene. None of the restriction maps agree closely with that of the RCA I cDNA, however. With the exception of pCBG3H1, the restriction maps of the clones diverge more from each other than do the ricin and RCA I maps.

IV.1.B. Nucleotide sequence diversity

DNA sequencing was carried out on the entire pCBG3H1 (lambda clone 10) ricin gene, most of the putative lectin gene from lambda clone 1 (subcloned into pCBG1R2 and pCBG1R5), a short region of the putative lectin gene in pCBG2H1 (lambda clone 3) and a short region of the putative lectin gene in pCBG4H2 (from lambda clone 7). An overview of the sequencing strategy is given in Figure 15, whilst details of the M13 constructs and sequencing primers used are shown in Tables 7 and 8 respectively.

IV.1.B.1) Nucleotide sequence of the pCBG3H1 (lambda clone 10) ricin gene

The nucleotide sequence of the pCBG3H1 (lambda clone 10) ricin gene and its flanking regions is shown in Figure 16. The coding sequence of the gene is shown in Figure 16.b whilst the sequences upstream and downstream from the coding region are shown separately in Figures 16.a and 16.c respectively. The most probable positions of the mRNA 5' and 3' termini, which were identified by RNase protection (Figures 22 and 23) are shown. On the basis of these data, the 5' untranslated region of the ricin gene appears to be 60 nucleotides in length, whilst the 3' untranslated region is calculated to be 158 nucleotides. The nucleotide sequence of the coding region confirms that pCBG3H1 (lambda clone 10) contains a ricin gene. In Figure 16.b, the pCBG3H1 sequence is compared with the nucleotide sequences of the pRCL617 ricin cDNA (Lamb et al., 1985) and pAKG ricin genomic (Halling et al., 1985) clones previously described. The pCBG3H1 sequence differs from that of pAKG at two nucleotide positions and from that of pRCL617 at only four. The pCBG3H1 (lambda clone 10) ricin gene must therefore encode a ricin D type polypeptide. Despite the small differences in their nucleotide sequences, the pCBG3H1, pAKG and pRCL617 genes each encode a polypeptide of the same primary sequence. As expected, the pCBG3H1 ricin gene coding region does not contain any introns. This appears to be a common feature of plant lectin genes (Vodkin et al., 1983, Kaminski et al., 1987). The flanking regions of the pCBG3H1 ricin gene

were found to share a high degree of homology with the corresponding regions in the pAKG clone. Within the 60 bp of 5' untranslated region and 250 bp of 5' flanking region sequenced (Figure 16.a), the pCBG3H1 sequence shows one insertion of 1 bp, five 1 bp deletions and two 1 bp substitutions with respect to the pAKG sequence. Two potential TATA boxes plus a possible CAAT/AGGA analogue previously identified in the pAKG ricin gene 5' flanking region (Halling et al., 1985) are also present in the pCBG3H1 ricin gene promoter. The promoter sequence is discussed in more detail in section IV.4. Within the 158 bp of 3' untranslated region and 163 bp of 3' flanking region sequenced in the pCBG3H1 clone (Figure 16.c), a total of three 1 bp deletions, one 1 bp insertion and three 1 bp substitutions were found with respect to the pAKG clone. The two potential poly (A) signals of sequences AATAAA and AATAAG which are found in all the castor bean lectin clones previously described (Roberts et al., 1985, Halling et al., 1985) are also conserved in the pCBG3H1 clone at 102 and 49 nucleotides upstream respectively from the most probable poly (A) site. The 3' sequence in Figure 16.c was scanned for two other elements thought to be involved in the polyadenylation of eukaryotic mRNAs. Firstly, no trace could be found of the YGTGTTY element (or G/T cluster) commonly present in mammalian genes downstream from the AATAAA sequence (McLaughlin et al., 1985). Secondly, no match could be

found for the CAYTG consensus element commonly found near the poly (A) site of animal genes (Berget, 1984).

IV.1.B.11) Partial nucleotide sequence of the lectin
pseudogene from lambda clone 1

DNA sequencing was carried out on two subclones from lambda clone 1, namely pCBG1R2 and pCBG1R5. The inserts in these two plasmids cover the 5' and 3' portions respectively of a putative lectin gene. As is shown in Figure 11, an Eco RI fragment with an expected size of 344 bp is presumed to link these two subcloned regions in the R. communis genome. However, since this fragment was not identified on the Southern blot in Figure 9, it was not subcloned. The nucleotide sequence data obtained, which is shown in Figure 17, demonstrates that three frameshift mutations are present in the putative lectin gene coding region. Thus the putative lectin gene in lambda clone 1 must be a non-functional pseudogene. The pseudogene shares homology with the pCBG3H1 and pAKG ricin genes from a point 19 bp upstream from the start of the ricin coding sequence (Figure 17.a) through most of the region sequenced. A striking repetitive element, consisting of the trinucleotide ATA repeated 18 times, is seen in the 5' flanking region of the pseudogene (Figure 17.a). The significance of this element is not known, but it may perhaps be associated with a DNA transposition event (P.

Gilmartin, pers. comm.). This might explain why the Nco I and Eco RV sites in the 5' flanking region of the lambda clone 1 pseudogene occur approximately 200 bp further upstream than do the same sites present in the pCBG2H1 (lambda clone 3) pseudogene 5' flanking region. Significant homologies with pRCL617 and pAKG could be found downstream from the "coding " region of the pseudogene, but a considerable number of deletions and insertions appear to have occurred in this area. Nevertheless the two potential poly(A) signals found in the ricin gene 3' untranslated region (Figure 16.c) are present. No significant homologies could be found with the 3' flanking regions of the pCBG3H1 and pAKG genes. In general, the nucleotide sequence of the lambda clone 1 (pCBG1R2/pCBG1R5) pseudogene differs more from the ricin and RCA I cDNA sequences than do the respective lectin cDNA sequences from each other. The most striking example of the sequence degeneration which has taken place, presumably since the gene became non-functional, is the 188 bp "scrambled" region in the pCBG1R2 insert (Figure 17.b). This region contains sequence which is unrecognisable as lectin coding sequence and could not be aligned with the lectin cDNA sequences. Moreover, the expected size of this region based on the cDNA sequences is 216 nucleotides. Thus even if this region were translated, the sequence could only be read as far as position 193, where a TAA translation termination codon

(highlighted in Figure 17.b within a box) occurs. Two other frameshift mutations were identified; one at position 647 in the A chain region (Figure 17.b) and another at position 55 in the B chain region (Figure 17.c). In order to establish the identity of the lectin pseudogene, its degree of homology with the respective lectin cDNAs (within the areas sequenced) was calculated. If all the "coding" sequence determined (ie that in figures 17.b and 17.c, apart from the "scrambled" region) is taken into account, the pseudogene shares 87.9% homology with the ricin cDNA and 86.3% homology with the RCA I cDNA at the nucleotide sequence level. Thus, the lectin pseudogene shares a marginally greater homology with the ricin cDNA. Figures 17.b and 17.c show nevertheless that many of the mismatches occur at positions where the ricin and RCAI cDNA sequences are the same. A more informative way of determining the identity of the lectin pseudogene is to compare its DNA sequence with those of the ricin and RCA I cDNAs only at "codons" where the two lectin primary sequences diverge. The amino acids specified at each of the relevant "codons" may be determined and compared with the corresponding codons in the ricin and RCA I cDNAs. This is somewhat artificial in that the pseudogene does not actually encode a protein, but the results are far more revealing. Out of a total of 48 codons of divergence, the lectin pseudogene agreed with the ricin cDNA at 32 positions and with the RCA I cDNA at

only 8 positions. At the remaining 8 positions, the lectin pseudogene was in agreement with neither. The pseudogene lacks a ricin-specific GCT/alanine codon at position 485 in the A chain region (Figure 17.b), so this was counted as an RCA I-like codon. Despite some affinities with the RCA I cDNA clone, it is apparent that the cloned pseudogene is essentially ricin-like. Interestingly, many of the RCA I-like codons were found at the 3' end of the B chain sequence (Figure 17.c). In this respect, the pseudogene therefore shows some similarities with the ricin E cDNA clone isolated by Ladin et al (1987). The fact that the lambda clone 1 pseudogene does not contain a poly(A) tract at its 3' end indicates that it is an unprocessed pseudogene.

IV.1.B.iii) Nucleotide sequence of an internal region from the putative lectin gene in pCBG2H1 (lambda clone 3).

The DNA insert in the plasmid pCBG2H1 was subcloned from lambda clone 3 and was found to contain a putative lectin gene. Although it was shown by RNase protection analysis that the putative lectin gene in pCBG2H1 is transcriptionally non-functional in vivo (see section IV.2), a short region of the presumed pseudogene was nevertheless sequenced to determine whether it is ricin-like or RCA I-like. The region in question occurs near

the 3' boundary of the A chain "coding" sequence (Figure 18). The fact that pCBG2H1 contains a pseudogene is confirmed by the presence of two frameshift mutations in the sequence; one at position 91 and one at position 165. Both are a result of a single nucleotide insertion. Interestingly, the second of these two mutations occurs at essentially the same position as a frameshift mutation identified in the lambda clone 1 pseudogene A chain sequence (Figure 17.b). Since the sequences of the two pseudogenes diverge slightly around this position, it is not possible to unequivocally identify which nucleotide originally caused the frameshift mutation. Nonetheless, it seems highly likely that the respective mutations in the two pseudogenes arose in a common ancestral gene or pseudogene. The restriction maps of the two pseudogenes show certain similarities which add weight to this hypothesis. Furthermore, a comparison between the sequence in Figure 18 and the corresponding sequence of pCBG1R2 clone (Figure 17.b), reveals that the two pseudogenes share 97.1% homology at the nucleotide sequence level within the region in question. In contrast, the pCBG2H1 pseudogene shares 87.1% and 85.9% homology respectively with the ricin and RCA I cDNAs within the same region. Again, in order to assess whether the pCBG2H1 pseudogene is ricin- or RCA I-like, the nucleotide sequence in Figure 18 was compared with the respective lectin cDNA sequences only at codons where the

ricin and RCA I primary sequences diverge. Out of a total of 7 codons of divergence, the pCBG2H1 pseudogene agreed with the ricin cDNA at 4 positions and with the RCA I cDNA at 3 positions. It was therefore not possible to classify this pseudogene as either ricin-like or RCA I-like.

IV.1.B.iv) Nucleotide sequence of an internal region from the putative lectin gene in pCBG4H2 (lambda clone 7)

The plasmid pCBG4H2 contains a 3.0 Kb DNA fragment subcloned from lambda clone 7 which was found to contain a putative lectin gene. RNAse protection analysis showed, as with the pCBG2H1 clone, that the putative lectin gene in pCBG4H2 is transcriptionally non-functional in vivo (see section IV.2). Nevertheless, in order to establish whether the pCBG4H2 pseudogene is ricin- or RCA I-like, a short internal region was sequenced. The 90 bp nucleotide sequence determined, which occurs towards the 5' end of the B chain region, is shown in Figure 19. No frameshift mutations occur within this region. The pCBG4H2 pseudogene shares 96.7% homology with the ricin cDNA and 80.0% homology with the RCA I cDNA in the region sequenced, suggesting that it is ricin-like. Similarly, out of a total of 8 codons where the ricin and RCA I genes specify different amino acids, the pCBG4H2 pseudogene sequence was in agreement with ricin at 7 positions. At

the 1 remaining codon, the pCBG4H2 pseudogene was in agreement with neither ricin nor RCA I. The available data suggests, therefore, that pCBG4H2 contains a ricin pseudogene. It is perhaps worth noting that the B chain regions of the pCBG4H2 pseudogene and ricin cDNA clones have virtually identical restriction maps (see Figure 15). The restriction maps of the respective A chain regions differ considerably, however.

To summarise, four different putative lectin genes were analysed by DNA sequencing. As expected, the nucleotide sequence of the lectin gene in pCBG3H1 (lambda clone 10) was virtually the same as the ricin cDNA and genomic sequences previously determined (Lamb *et al*, 1985, Halling *et al*, 1985). None of the clones appeared to contain an RCA I gene, although some similarities with the RCA I cDNA sequence were observed in both the pCBG1R2/pCBG1R5 (lambda clone 1) and pCBG2H1 (lambda clone 3) clones. The putative lectin gene in pCBG4H2 (lambda clone 7) appeared to be essentially ricin-like. The presence of frameshift mutations in the pCBG1R2/pCBG1R5 (lambda clone 1) and pCBG2H1 (lambda clone 3) sequences demonstrated that they contain non-functional pseudogenes, as was subsequently shown to be the case with the pCBG4H2 (lambda clone 7) clone. It is probable that much of the nucleotide sequence and restriction site diversity of the pseudogenes has arisen since these structures became non-functional.

IV.1.C. Size and quality of genomic libraries

IV.1.C.1) Estimated size of a "complete" genomic library

The castor bean DNA samples used in this study for genomic cloning and Southern blotting were extracted from preparations of isolated nuclei (section II.3). Nuclei yields were quantified prior to DNA extraction, as were the DNA yields in the final preparations. Thus an approximate estimate of the R. communis haploid genome size or C-value could be calculated in each case.

Although the R. communis seedlings used were of unknown variety, it was assumed that the cultivar is diploid, as has been shown for other cultivars (D. Griffiths, pers. comm.). Table 4 shows the individual and mean values calculated. It can be readily seen that the calculated C-values show considerable variation around the mean value. Thus, within 95 % confidence limits, the estimated haploid genome size of R. communis is $0.26 \text{ pg} \pm 0.17 \text{ pg}$, or $2.8 \times 10^8 \text{ bp} \pm 1.8 \times 10^8 \text{ bp}$. The corresponding number of recombinant clones in a "complete" (ie 99% representative R. communis genomic library is therefore calculated, using the Clarke and Carbon (1976) formula, as $8.6 \times 10^4 \pm 5.6 \times 10^4$ clones, assuming an average DNA insert size of 15 Kb.

Three main factors determine the accuracy of the C-values obtained, namely the accuracy of the nuclei count, the efficiency of DNA recovery from the nuclei and the accuracy of DNA estimation. Although the accuracy of the DNA estimation is high, significant random error may have occurred with the nuclei counts. Furthermore, the efficiency of DNA recovery from the nuclei is probably subject to substantial variation due to the experimental procedures used. The high standard error in the results is therefore expected. During the DNA extraction process, a dialysis step is carried out which undoubtedly results in the loss of an appreciable amount of the DNA. The actual haploid genome size of R. communis must therefore be significantly greater than the mean value calculated. The underestimation inherent in the mean C-value is not accounted for by the calculated standard error.

IV.1.C.ii) Quantity and quality of recombinant bacteriophages in the genomic libraries

Table 5 shows the respective numbers of bacteriophages obtained in the construction of the partial Eco RI and partial Sau 3A genomic libraries. The yield of bacteriophages obtained from each in vitro packaging reaction is shown, as is the corresponding cloning efficiency calculated in each case. Only one in vitro packaging reaction (reaction 1) was carried out to obtain

the partial Eco RI library, but two separate reactions (reactions 2 and 3) were used to obtain the partial Sau 3A library. It is immediately apparent that a vastly greater yield of bacteriophages was obtained in the preparation of the partial Sau 3A library. Unfortunately, none of the ligated DNA samples were analysed by gel electrophoresis, so the differences in bacteriophage yields observed cannot be unequivocally attributed to either differences in ligation efficiency or differences in packaging efficiency. However, the in vitro packaging mixture used in reaction 1 had been in storage for a far greater time than those used in reactions 2 and 3 and it seems likely that appreciable deterioration had taken place by the time it was used. Another contributory factor is the DNA concentration in the ligation reactions. Samples 2 and 3 were ligated at a DNA concentration twice that of sample 1. The formation of lambda DNA concatamers, which provide the best configuration for in vitro packaging, is favoured at high DNA concentrations. Presumably the difference in bacteriophage yields between reactions 2 and 3 may be explained by their different vector: insert stoichiometries. The vector insert ratio used in reaction 2, which produced the greatest yield of bacteriophages, is the same as the optimum predicted theoretically (Maniatis et al. 1982). Nevertheless it is common practice to set up ligations using different vector: insert ratios, since the optimum ratio may sometimes deviate in practice from that

expected. The most likely reason for such deviations is that a significant number of the molecules in some substrate DNA preparations may lack cohesive termini. This may be caused by factors such as exonuclease degradation.

When assessing whether or not the genomic libraries constructed are "complete" it must be borne in mind that some of the bacteriophages obtained are non-recombinants. This is unavoidable, since it is impossible to remove all traces of the stuffer fragments from the lambda arms preparation and the charon 35 vector used does not provide any means for the selection of recombinants. If desired, the percentage of non-recombinants in the in vitro packaged bacteriophages could be determined by carrying out plaque hybridisations using a charon 35 stuffer fragment probe. This is normally considered unnecessary, however. Previous estimates suggest that the proportion of non-recombinants in the genomic libraries prepared for this study should be no greater than 5-10% (Murray et al. 1984). Even if the libraries are assumed to contain 10% non-recombinants and the upper estimate of the R. communis genome size (4.6×10^8 bp) is used, a "complete" library (Clarke and Carbon, 1974) would still be represented by only 1.8×10^8 bacteriophages with inserts of average size 14 Kb in the recombinants. This figure is still probably an underestimate due to the error inherent in the C-value

estimation, as discussed earlier. Nevertheless, it appears that the 9×10^5 bacteriophages screened from the partial Sau 3A library contained approximately 35-40 positive clones roughly equivalent to 4 genomes. This suggests that a "complete" R. communis genomic library would be represented by only $2-3 \times 10^5$ of the same bacteriophages.

On the basis of the figure above, it is clear that the partial Eco RI library constructed was not "complete". The fact that only 2 positive clones were identified amongst the recombinants confirms this conclusion, and suggests further that the library contains a significantly non-random complement of genomic clones. This is not unexpected, since Eco RI recognises a hexanucleotide sequence which inevitably occurs non-randomly within the R. communis genome. In contrast, the restriction enzyme Sau 3A recognises a tetranucleotide sequence which occurs far more frequently in genomic DNA. Sau 3A may therefore be used to produce libraries which may be considered to be essentially random (Maniatis *et al*, 1982). Additional precautions were also taken to ensure that the partial Sau 3A library would be as random as possible (section II..B).

In summary, it appears that the partial Eco RI library constructed was both incomplete and significantly non-random, whilst the bacteriophages screened from the

partial Sau 3A library contained approximately 4 genomes of essentially randomly cloned DNA.

IV.1.C.iii) Possible reasons for the uneven distribution of lectin clone types amongst the 17 clones isolated

Although 5 different members of the R. communis lectin gene family have been cloned and characterised, it is clear that at least one member of the family remains to be isolated, since no RCA I genomic clones were obtained. On the basis of the Southern blot (Figure 5), it was estimated that the R. communis gene family contains approximately 8 members, so the number of lectin genes which remain to be isolated may be as many as three. Three factors determine the degree of randomness of the pool of lectin-positive lambda clones obtained. The first and most obvious factor is the degree of randomness of the genomic libraries screened. A second factor is the degree of randomness involved in selecting putative positive clones for rescreening;. In the case of the 9×10^5 bacteriophages screened from the partial Sau 3A library, only a third to a half of the potential positives identified in the first round of screening were rescreened. The third factor which must influence the degree of randomness of the pool of lectin-positive clones is sampling error. In a statistically ideal situation,

the complement of lectin-positive clones would be expected to consist of approximately 2 clones of each member of the gene family. It can be seen, however, that the 17 clones isolated represent only a small sample, so the absence of certain genes from the pool of clones might be explained merely by sampling error alone.

The lack of RCA I genomic clones is probably attributable either to non-random selection of genomic clones and/or sampling error. Although care was taken to select a range of lectin-positive clones producing hybridisation signals of different intensities, it appears that there has been some selection bias towards the more strongly hybridising clones (Table 6). Nevertheless, many of the clones isolated appear to be no more homologous with the ricin cDNA probe than is the RCA I cDNA (see Figure 10). The strongest evidence that the selection of putative positive clones for rescreening was non-random is shown by the fact that 10 of the 17 lectin-positive clones fall into group C (Table 6), and therefore appear to contain ricin-like genes. 3 of the clones (clones 6, 12 and 16), however, grew poorly and therefore produced only weak hybridisation signals when screened, so they cannot have been chosen in preference to the clones in the other groups. A second possible reason for the large number of group C clones isolated is that more than one authentic ricin gene might be present in the R. communis genome. This conclusion is

in agreement with the high stringency genomic blot shown in Figure 7, which suggests that the R. communis genome contains 2 ricin-like genes. It is possible, therefore, that some of the lambda clones in group C might contain variant ricin genes.

In summary, it appears that the absence of any RCA I genes in the 17 lectin-positive clones may be attributed to statistical sampling error and/or non random selection of positives. The large number of clones in group C may indicate, apart from some selection bias, the presence of more than one authentic ricin gene in the R. communis genome.

SECTION IV.2. IN VIVO TRANSCRIPTION OF CASTOR BEAN LECTIN
GENES

IV.2.A. Northern analysis of transcripts

Total RNA was isolated from castor bean seeds of a variety of developmental stages, from the endosperm of germinating seeds and from leaf and root tissues (section III.5.A). In order to assess the overall developmental and tissue-specific pattern of lectin gene transcription, a gel containing samples of the RNA preparations was Northern blotted onto a nylon membrane which was probed with a ricin cDNA and washed at low stringency (section III.5.B). Both ricin and RCA I transcripts will be detected under such conditions. The resulting autoradiograph, shown in Figure 21, clearly demonstrates that lectin mRNAs are synthesised in a seed-specific fashion. No lectin mRNA is detected in the pre-testa stages of seed-development, but by stage D, a visible band is evident. The level of lectin transcripts increases gradually up to and including stage F, but no lectin mRNA is detected in the dry seed tissue. Similarly, lectin transcripts are not detected in either leaf or root tissue or in the endosperm of germinating seeds.

The quality of the RNA used, at least in the preparations obtained from seed developmental stages D to F, appears to be very good, as judged by the sharpness of the bands on the autoradiograph in Figure 21. The fact that only a single band may be seen in the relevant tracks confirms that the ricin and RCA I transcripts are essentially the same size. The observed migration of the lectin mRNA species agrees with the sizes of the lectin cDNA clones. Since negligible RNA degradation could be seen, the RNA preparations were judged to be of sufficient quality for use in RNase protection experiments.

IV.2.B. Confirmation of transcriptional inactivity of pseudogenes in pCBG2H1 (lambda clone 3) and pCBG4H2 (lambda clone 7)

DNA sequencing and RNase protection analysis were carried out in parallel on the putative lectin genes in pCBG2H1 (lambda clone 3) and pCBG4H2 (lambda clone 7). As has already been discussed in section IV.1.B.iii), the DNA sequence of an internal region from the putative lectin gene in pCBG2H1 revealed the presence of frameshift mutations. Thus pCBG2H1 (and therefore lambda clone 3) must contain a non-functional pseudogene. The RNase protection data obtained confirms this conclusion and also shows that the putative lectin gene in pCBG4H2 (lambda clone 7) is transcriptionally non-functional.

IV.2.B.1) RNAse protection analysis of transcripts using an internal probe from the putative lectin gene in pCBG2H1 (lambda clone 3)

In order to assess whether the putative lectin gene in pCBG2H1 (lambda clone 3) is transcribed in vivo, an RNAse protection experiment was carried out. An internal probe extending from the 3' end of the A chain region into the 5' end of the B chain region (see Figure 15 and Table 9) was used so that any transcripts sharing 100% homology with the clone could be easily identified by their size. Figure 24 shows the result obtained.

If the putative lectin gene in pCBG2H1 (lambda clone 3) were expressed in vivo, a protected fragment of the same size as the 324 nucleotide cloned insert in the pGEM-blue3JT3 plasmid would be expected. No such protected fragment was observed, however, even when the gel was exposed for a much longer period of time. The only RNAse protected fragments observed were much smaller, the maximum size being less than 90 nucleotides. Thus whilst some of the in vivo transcripts must share partial homology with the cloned sequence, it was concluded that the lectin gene-like structure in pCBG2H1 (lambda clone 3) is non-functional.

IV 2.B.ii) RNAse protection analysis of transcripts
using an internal probe from the putative
lectin gene in pCBG4H2 (lambda clone 7)

A similar RNAse protection experiment to that described above was carried out in order to assess whether the putative lectin gene in pCBG4H2 (lambda clone 7) is transcriptionally active in vivo. In this case an internal probe spanning a 380 nucleotide region in the A chain sequence (see Table 9 and Figure 15) was used. No RNAse protected species in the size range of the markers used (67 nucleotides and above) could be identified on the autoradiograph obtained (data not shown). Thus pCBG4H2 (and therefore lambda clone 7) must contain a transcriptionally non-functional pseudogene.

To conclude, a total of three different lectin pseudogenes, present in lambda clones 1,3 and 7 respectively, have been identified. The pCBG2H1 (lambda clone 3) and pCBG4H2 (lambda clone 7) pseudogenes have been shown to be transcriptionally inactive in vivo, whilst DNA sequencing has shown that the lambda clone 1 and lambda clone 3 pseudogenes contain frameshift mutations. Therefore the R. communis lectin gene family appears to have a high proportion of non-functional members. This is not, however, unusual. For example, a similar situation occurs with the lectin gene family of

pea which consists of 4 different members, only one of which is functional (Kaminski et al., 1987).

IV.2.C. Transcription of the pCBG3H1 (lambda clone 10)
ricin gene in vivo

In view of the high level of homology which the pCBG3H1 (lambda clone 10) ricin gene shares with the ricin cDNA and genomic clones previously described (Lamb et al., 1985, Halling et al., 1985), it was considered highly likely that this gene is expressed in vivo. RNase protection analysis was therefore used to verify that the gene is transcribed in vivo, to map the 5' and 3' termini of its mRNA and to characterise its tissue-specific and developmental pattern of expression. Two RNA probes were prepared by in vitro transcription for this purpose. A probe spanning the 5' end of the gene was used to identify potential cap sites, whilst a probe spanning the 3' end of the gene was used to identify potential poly(A) sites. At the same time, the developmental and tissue-specific expression pattern of the gene could be monitored. Figure 22 shows the result of the RNase protection experiment carried out using the 5' terminal probe whilst Figure 23 shows the corresponding result obtained with the 3' terminal probe.

IV.2.C.1) Tissue and developmental specificity of
transcripts

A sizeable number of RNase protected species are evident on the autoradiograph shown in Figure 22. Nevertheless the autoradiograph shows, as did the Northern blot (Figure 21), that the accumulation of lectin transcripts occurs in stages D, E and F of seed development. Again, a gradual increase in the levels of these transcripts is observed during this period. The developmental profile of lectin mRNA accumulation seen in Figure 22 is comparable with the corresponding patterns observed at the protein level (Gifford et al, 1982) and at the level of translatable mRNA (Roberts and Lord, 1981A.). Thus, the expression of lectin genes in R. communis may be regulated, at least in part, at the transcriptional level. Despite the overall changes in mRNA levels seen during the period of seed development, the relative abundance of the RNase protected species appears to be the same in each sample. The significance of the individual bands seed will be discussed in the following section. No lectin transcript bands were observed in any tracks other than those corresponding to seed stages D, E and F when the gel was autoradiographed for a short period of time. Upon longer exposure however, the pattern of bands observed in stages D to F was also seen in the tracks corresponding to stage C seed tissue, dry seed tissue and germinating seed endosperm tissue harvested both 3 and 7 days after imbibition. The unevenness of the bands seen with the seed stage C and dry seed samples suggested that

they may have arisen from a slight leakage of the gel loading wells. There can be no doubt however, that trace levels of lectin transcripts are present in the endosperm tissue of germinating seeds both 3 and 7 days after imbibition. The level of lectin transcripts in the 7 day sample is in fact greater than that in the 3 day sample. This result is very surprising, since it might imply that lectin gene transcripts are synthesised de novo during germination (see below).

Figure 24 shows the result of the corresponding RNase protection experiment carried out using the 3' terminal ricin gene probe. The developmental pattern of lectin gene expression seen in this autoradiograph appears to be similar to that observed in Figure 22. In this case however, the gradation of hybridisation signals seen between the tracks corresponding to seed stages D to F is less marked. This is due in part to the fact that the autoradiograph was exposed for longer than that shown in Figure 22 and printed at lower contrast, so that the fainter bands could be seen more clearly. On shorter exposures of the same gel, the gradation of signal intensities in the tracks corresponding to seed stages D to F resembled more closely that seen in Figure 22. Nevertheless it appears that the amount of RNA removed from the seed stage D preparation for analysis may for some reason have been greater than the expected 10 ug,

since many of the larger bands in track 6 (seed stage D) are of at least the same intensity as their counterparts in track 7 (seed stage E). The presence of RNase protected species in the tracks corresponding to the germinating seed endosperm and dry seed tissues can be clearly seen in Figure 23. Faint bands of the same pattern could also be seen in the stage C track. No well leakage was observed when the gel was loaded, so trace levels of lectin mRNA must presumably appear during seed development by stage C. Similarly, a residual level of lectin mRNA must apparently remain in the seed after desiccation has taken place. Although the level of lectin mRNA is even lower in the seed endosperm after 3 days of germination, the autoradiograph in Figure 23 clearly demonstrates that the lectin mRNA level has increased after 7 days of germination. This confirms the observation made when the 5' terminal probe was used and suggests that lectin transcripts might be synthesised de novo at low levels during germination. Such an occurrence would be very surprising, as the castor bean lectins are being broken down to be used as a nitrogen source for the seedling at this time. The synthesis of more storage proteins would therefore be counter-productive and would require energy which could only be gained from the reserves stored in the seeds. Furthermore, if the lectin mRNAs seen were translated, the endosperm of the germinating seeds might not be able to segregate the ricin

polypeptides produced. An alternative explanation for the apparent increase in lectin mRNA levels seen is that the abundance of these molecules may increase only relative to the overall level of RNA in the absence of any de novo synthesis. Since the endosperm tissue gradually degenerates during germination, the ribosomal and transfer RNAs must eventually be broken down and the products translocated to the growing parts of the embryo. However, it is already well-established that a number of proteins, notably glyoxysomal proteins, are synthesised on the 80S ribosomes of the castor bean seed endosperm at this stage during germination (Bowden and Lord, 1976). Thus neither of the two possible explanations for the increase in lectin mRNA levels seen by day 7 can be confirmed at this stage.

In summary, the transcription of lectin genes in R. communis is seed-specific. Trace levels of lectin mRNAs are detectable by the late pre-testa stage (stage C), but lectin gene transcription appears to occur mostly in post-testa seeds(stage D to stage F). Only very low levels of lectin mRNA are detectable in the dry seed, but these levels persist in the seed endosperm during germination, the relative abundance of the lectin mRNAs apparently increasing by the seventh day of germination. No transcription of the lectin genes appears to occur in the root or leaf tissues.

IV.2.C.ii) Potential cap sites of the pCBG3H1 (lambda
clone 10) ricin gene

Unfortunately the identification of the cap site(s) of pCBG3H1 (lambda clone 10) ricin gene is considerably complicated by the large number of RNase protected species seen in the autoradiograph in Figure 22. Although multiple RNase protected bands will be detected from a single gene if transcription is initiated at more than one position, the total number of bands seen in each track is too large to be accounted for by one gene alone. It can be readily seen on the autoradiograph in Figure 22 that the principal RNase protected species occur as a series of doublets. Each doublet is attributable to two fragments of similar abundance differing in size by only nucleotide. In most cases, a third, fainter band arising from a fragment 1 nucleotide smaller is also present below the doublet. There is little doubt that this doublet/triplet pattern is an artefact. In common with most other plant genes, the ricin gene contains a high percentage of adenine and thymine residues at its 5' end and in the region of the cap site. When nuclease protection analysis is performed, this may frequently cause a phenomenon known as "breathing", whereby the complementary RNA strands are able to shift slightly from their optimum alignment. During the digestion process, this may result in RNase-protected species of smaller size than that expected being

generated. This process is frequently referred to as "nibbling" (Williams and Mason, 1985). When interpreting the result of a nuclease protection experiment, the possibility of "nibbling" must be taken into account, and there can be little doubt here that "nibbling" has occurred. The RNase-protected species of interest are therefore those corresponding to the uppermost band in each doublet/triplet. A total of 7 principal bands may be identified in this way. Although some plant genes have been observed to initiate transcription from more than one position (eg Herrman et al., 1988), it is unlikely that the 7 bands are all attributable to a single gene. Thus it appears that a number of lectin genes are transcribed in vivo in R. communis. Considering the size of the lectin gene family, the bands seen on the autoradiograph in Figure 22 may well be attributable to both ricin and RCA I genes. The size of each RNase-protected fragment can be accurately determined, since the bands are numerous and regularly spaced. Using this information, a series of potential cap site locations may be predicted. Figure 25 shows the positions of these potential cap sites. Two of the seven sites are of particular interest and have been labelled separately. Firstly, the potential cap site position corresponding to the most intense band seen in Figure 22 is indicated. Although this position does not necessarily correspond to an actual cap site, the fact that it occurs 16 nucleotides upstream from that proposed

Figure 25



Translation initiation codon

-15
 TIGCTGCAATCAAAGATG

- ▼ - cap site position predicted from the size of the largest RNase protected fragment
- ▼ - cap site positions predicted from the sizes of the other main RNase protected fragments
- ▲ - potential cap site position corresponding to the most intense band
- ▼ - cap site positions predicted for pARG ricin genomic clone (Halling et al, 1985)

Figure 25 Predicted cap site(s) of the pCBG3H1
(lambda clone 10) ricin gene

The positions of a number of potential cap sites identified upstream from the pCBG3H1 (lambda clone 10) ricin gene coding sequence are shown. The proposed cap site positions were determined from the results of an RNase protection experiment described in Figure 22. The potential cap site positions of the pCBG3H1 ricin gene are compared with those determined by Halling et al (1985) for a similar ricin genomic clone, pAKG. Nucleotide coordinates shown are relative to the position of the ATG translation initiation codon.

by Halling et al (1985) for the pAKG ricin gene suggests that it might. Nonetheless three other potential cap sites occur further upstream. It is highly likely that the potential cap site located at -60 (with respect to the ATG translation initiation codon) corresponds to an actual cap site, since the transcript which gave rise to the protected fragment in question must share the greatest homology with the probe used. If the pCBG3H1 (lambda clone 10) ricin gene uses the cap site at -60 exclusively in vivo, then it is presumably transcribed at only a low level, as judged by the faintness of the corresponding band on the autoradiograph (Figure 22). Nevertheless it is possible that some of the other bands observed might be attributable to the same ricin gene if multiple cap sites are used. A number of plant genes have been found to initiate transcription from more than one position (eg Herrman et al, 1988). There may possibly be some significance in the fact that the pCBG3H1 ricin gene 5' flanking region appears contains more than one potential TATA box (Figure 16.a). Another factor which must be borne in mind is that the cap site does not necessarily correspond to the transcription initiation site, since the 5' processing of mRNAs may involve the removal of several nucleotides (Grierson, 1982). It is therefore not impossible that multiple size mRNA species could arise through heterogeneous processing of the mRNA 5' ends.

Assuming that the potential cap site at -60 is genuine, the 5' untranslated region of the mature mRNA transcribed from pCBG3H1 ricin gene is apparently 25 nucleotides longer than that proposed by Halling et al (1985) for the pAKG ricin gene, but approximately 17 nucleotides shorter than that predicted by Lamb (1984) from primer extension studies using a ricin cDNA. The general dissimilarity between these respective data suggests that transcription is initiated at diverse positions in different ricin genes. One of the potential cap sites identified in the pCBG3H1 ricin gene 5' flanking region does in fact coincide with a potential cap site identified by Halling et al (1985) in the pAKG ricin gene 5' flanking region. In the latter case, however, the putative cap site position was predicted from the size of an S1 nuclease protected fragment which may be an artefact caused by "breathing". Considering the close homology between the pCBG3H1 and pAKG ricin genes, it is surprising that the pattern of fragments seen in Figure 22 is almost completely different to that observed by Halling et al (1985). Presumably the difference in results is attributable to nucleotide sequence divergence between the members of the lectin gene family or possibly even nucleotide sequence divergence between the corresponding lectin genes of the two varieties of R. communis used.

Although it is not easy to determine which of the seven potential cap site positions correspond to actual cap sites actually used in vivo, it seems highly unlikely that all the positions identified are used by the pCBG3H1 ricin gene. In particular the sites identified at -23 and -28 are located at an appreciable distance downstream from the more 3' TATA sequence, suggesting that they might not be used in vivo. One possible explanation for the result seen in Figure 22 is that the most 5' cap site identified might be associated with the more 5' TATA sequence, which would therefore appear to function at only a low level. One or more of the remaining cap site positions may be associated with the more 3' TATA sequence, which might therefore act as a more efficient binding site for RNA polymerase II.

IV.2.C.iii) Potential poly (A) sites of pCBG3H1 (lambda clone 10) ricin gene

The number of different RNase protected species visible on the autoradiograph in Figure 23 is particularly large, even when compared with the result obtained with the 5' terminal ricin gene probe (Figure 22). Again some of the RNase-protected species observed may be caused by "breathing" of the RNA: RNA hybrids. Nevertheless, it is clear that the transcripts of a number of different genes are represented in the autoradiograph. Since some plant

genes have been shown to produce several mRNA species in vivo which are polyadenylated at different positions (Dean et al, 1986), it is likely that the same phenomenon might occur with the castor bean lectin genes. Thus the interpretation of the pattern of bands observed is not straightforward. Although it is probable that some artefactual bands are present, it is unfortunate that they cannot be readily distinguished in this case. For the purpose of identifying potential poly (A) sites in the pCBG3H1 sequence, only the four largest RNase-protected species were considered. These four species occur as two doublets, each consisting of two bands of similar intensity. The bands in the upper doublet are much fainter than those in the lower doublet. The size of each of the four RNase protected fragments was estimated and a corresponding series of poly (A) site locations were determined.

Figure 26 shows the positions of the potential poly(A) sites, plus those deduced by Lamb et al (1985) and Halling et al (1985) from the nucleotide sequences of three different ricin cDNA clones. The potential poly (A) site at position 162 is presumably used in vivo by the pCBG3H1 ricin gene, since the transcript which gave rise to the protected fragment in question must share the greatest homology with the probe used. Thus at least one of the transcripts produced from this gene must contain a 162

Figure 26



- ▽ - position of potential poly(A) sites predicted from estimated sizes of bands in upper doublet (Figure 23)
- ▽ - position of potential poly(A) sites predicted from estimated sizes of bands in lower doublet (Figure 23)
- ▽ - poly(A) site in ricin cDNA clone reported by Lamb et al (1985)
- ▽ - poly(A) sites in 2 ricin cDNA clones reported by Halling et al (1985)

Figure 26 Predicted polyadenylation site(s) of the
pCBG3H1 (lambda clone 10) ricin gene

The position of a number of potential poly(A) sites identified downstream from the pCBG3H1 (lambda clone 10) ricin gene coding sequence are shown. The proposed poly(A) site positions were determined from the results of an RNase protection experiment described in Figure 23. The potential poly(A) site positions of the pCBG3H1 ricin gene are compared with those deduced by Lamb et al (1985) and Helling et al (1985) from the nucleotide sequences of three different ricin cDNA clones. The nucleotide coordinates shown are relative to the TGA translation termination codon at the end of the ricin B chain coding sequence.

nucleotide 3' untranslated region. This mRNA appears to be produced at only a low level, however, as judged by the faintness of the corresponding band on the autoradiograph in Figure 24. Thus, if the pCBG3H1 (lambda clone 10) ricin gene uses the poly (A) site at position 162 exclusively in vivo, then it must presumably be expressed at only a low level. It is quite possible, however, that one or more of the other three potential poly (A) sites shown in Figure 26 might also be used by the gene. The position of one of these three poly (A) sites does in fact correspond to that of a poly (A) site previously identified in a ricin cDNA clone (Halling et al, 1985). The use of multiple poly (A) sites has been shown to occur with a number of plant genes (Dean et al, 1986).

SECTION VI.3. POSSIBLE TRANSCRIPTIONAL REGULATORY SIGNALS
IN THE pCBG3H1 (LAMBDA CLONE 10) RICIN GENE
PROMOTER

The nucleotide sequence of the 5' flanking region and 5' untranslated region of the pCBG3H1 (lambda clone 10) ricin gene is shown in Figure 16.a. Since the most likely cap site position was calculated to be 60 nucleotides upstream from the translation initiation codon, it is assumed that the 310 bp sequence shown includes 250 bp of the promoter sequence. In view of the RNase protection data obtained (Figure 22), it is possible that transcription might be initiated at more than one position, but this cannot be proved at present. In figure 16.a, the pCBG3H1 promoter sequence is compared with that of the pAKG ricin gene (Halling et al., 1985). The various sequence elements of interest are highlighted as boxes.

IV.3.A. Identification of TATA box(es)

It can be seen in Figure 16.a that a TATA sequence (box I) occurs 13 bp upstream from the proposed cap site. This element may well represent the classical eukaryotic TATA or Goldberg Hogness box, although its position is slightly closer to the cap site than is normally observed. A similar TATATA sequence (box II) is present at -50 (with

respect to the cap site). Although further upstream, the location of this element falls just within the range of TATA box positions determined for other plant genes. Interestingly, the first TATA sequence mentioned occurs within an 18 bp sequence (box V) which is almost perfectly repeated at -103. As was discussed in section IV.2.C.ii), the presence of more than one functional TATA box in the pCBG3H1 ricin gene promoter is not inconsistent with the RNase protection data obtained (Figure 22). It therefore seems possible that the most 5' cap site position determined (Figure 25) might be associated with the more 5' of the two putative TATA boxes, whilst one or more of the remaining cap site positions might be associated with the second putative TATA box. Interestingly, the second putative TATA box occurs within a 6 bp sequence which is present at a similar location in both the pCBG3H1 ricin gene promoter and the promoter of a castor bean 2S albumin gene (S.D. Irwin, unpublished data). The region in question is identified in Figure 27, which shows a comparison made between the promoter sequences of the respective genes. In view of this homology, it seems highly likely that the more 3' TATA sequence might function as an RNA polymerase II binding site in vivo

IV.3.B. Identification of a putative CAAT/AGGA box analogue

Figure 27

pCBG3M1 1 CCTAAAAAAGTAAATTACTCTAATCGACATTATGAAATTTAACTAATTCGGTTCTA
gcopia 1 ATTATCAATGTACATAAAAAATAAATAAATT-TAAAAATAATAGTTCAACAACATCAA

pCBG3M1 61 ATTATATAATTATTGGTTAAACCA-ATCAATTCGGTTAAACACTGCTATGCCATATCT
gcopia 60 GCGHAAAATCGATAAAAAAGATCAATTCACCAACAGAACT-TGC--ATCGTGGTCCC

pCBG3M1 120 GTCTCAATTTATATATGCTATGCACTTCGGTATTAAATTTATAAGTTATTTTATGATC
gcopia 117 ATACGTGTAAAGCTTAAAGCAGACCTCAAAACCATGGTAGCCACCCCATAAAAATGCCAG

pCBG3M1 180 AATGATTGTGCTTTTCCTT-ATATAAAAAATCTATTAGTGTTTTCTGTATTAAATT
gcopia 177 AATGACTGGCCATTCTATACATCTTACACCTCTCGACATGCACTCCCTAAATTCCT

pCBG3M1 239 ATAACTTCATCTTATGAGAAATGCTAATGTAATTTGGACAGCCAATAAAATTCAGAAATTG
gcopia 237 ATAAAGGCC-CATTACCATCAT-CTACTATATCCATAAGAGCATC---TTCTGATATT

pCBG3M1 299 CTGCAATCAAG
gcopia 292 CACTAA-CAATC] Start of coding sequences

Figure 27 Comparison of the pCBG3H1 (lambda clone 10)
ricin gene and castor bean 2S albumin gene
promoter sequences

The diagram opposite shows a comparison made between the promoter sequence of the pCBG3H1 (lambda clone 10) ricin gene and that of a similarly regulated castor bean 2S albumin gene isolated from the same genomic library. The unpublished 2S albumin gene promoter sequence, labelled gcepla, is shown with the kind permission of S.D. Irwin. The sequence alignment shown was achieved using the Beckman Microgenie system. Nucleotide matches are indicated by vertical dashes between the two sequences. The position of the putative ricin gene mRNA cap site is shown by means of an arrowhead. Boxes I and II show the positions of two potential TATA boxes identified in the ricin gene promoter region whilst box III shows the position in the ricin gene promoter of a putative CAAT/AGGA box analogue. Box IV shows the position in the ricin gene promoter of a region sharing homology with a conserved element previously identified in the promoters of a number of seed protein genes (Dickinson et al, 1988). A region in the gcepla sequence, shown in box V, shares a significant homology with the box IV sequence, but the computer program used did not allow the alignment of the respective regions.

78 nucleotides upstream from the proposed cap site of the pCBG3H1 ricin gene is a CAAGT sequence (in box III) which Halling *et al* (1985) suggested might function as a CAAT box. It is possible, however, that the sequence ATTGA also included in box III might function as an AGGA box analogue, since it consists of a trinucleotide which mismatches the consensus sequence (G/T)NG and is flanked by adenine residues. Normally, however, two to five adenine residues are seen to occur on each side of the trinucleotide sequence. It can be seen in Figure 27 that there is no apparent conservation of the box III sequence in the 2S albumin gene promoter.

IV.3.C. Identification of DNA sequences which might confer seed-specific expression

Since the ricin and 2S albumin genes of *R. communis* are similarly regulated in an endosperm-specific fashion, it was hoped that the promoter sequence comparison shown in Figure 27 would reveal the presence of conserved sequences which might be involved in the seed-specific expression of the two genes. In the optimum alignment shown (which was achieved by means of the Beckman Microgenie system) however, only a putative TATA box region can be recognised as a significantly conserved sequence. A more striking sequence conservation was observed when the promoter sequences were compared by eye. A 14 bp region present at

-113 (box IV in Figure 27) shares a high degree of homology with a 16 bp region which occurs 110 bp upstream from the 2S albumin gene (box IV in Figure 27). A comparison between the respective conserved sequences is shown in Figure 28. The pCBG3H1 element is two nucleotides shorter than that in the 2S albumin gene promoter, but the two sequences agree at 13 out of 14 of the remaining positions. This high degree of homology seems to indicate that these elements contain sequences involved in the endosperm-specific expression of the two genes. Interestingly, both elements contain a sequence resembling the CATGCATG RY repeat found in the promoter region of a large number of seed protein genes (Dickinson *et al.*, 1988). The sequences of the ricin and 2S albumin gene elements differ from the classical sequence at only 1 and 2 positions respectively. Examples of seed protein genes which contain RY repeat elements in their promoters include two Phaseolus lectin genes (Hoffman and Donaldson, 1985), various pea legumin genes (Lycett *et al.*, 1985), a pea vicilin gene (Doyle *et al.*, 1986), a number of Arabidopsis 2S albumin genes (Krebbars *et al.*, 1988) and various cereal seed protein genes, which contain the RY repeat at or near the -300 element (Forde *et al.* 1985). Some promoters contain multiple copies of the RY repeat, such as the Phaseolus lectin genes, which were found to contain 3 copies (Hoffman and Donaldson, 1985). Similarly, the position of these elements has been found

Figure 28

	-112	-99
pCBG3H1	CATGCATCTT-C	CGT
gcep1a	CATACATCTTACACGT	
	-110	-95

Figure 28 Identification of a sequence motif conserved in the promoters of the pCBG3H1 (lambda clone 10) ricin gene and gcepla 2S albumin gene

A 14 bp region in the pCBG3H1 (lambda clone 10) ricin gene promoter was found to share a high degree of homology with a 16 bp region in the promoter of the castor bean gcepla 2S albumin gene. The respective sequences of interest have been aligned to achieve maximum homology. As is discussed in section IV.3.C, both sequences contain an RY repeat-like element (Dickinson *et al.*, 1988) which may be involved in determining their seed-specific expression patterns. Nucleotide coordinates shown above the pCBG3H1 sequence are relative to the most likely cap site position determined (Figure 25), whilst the nucleotide coordinates underneath the gcepla sequence are relative to the translation initiation codon of the gene (the cap site position of this gene has not yet been determined).

to vary considerably from 7 to approximately 300 nucleotides upstream from the cap site (Dickinson et al. 1988). The pCBG3H1 promoter does in fact contain, in addition to the RY repeat element described above, another similar element of the sequence TATGCATA further upstream at -143. Whilst this region is conserved between the pCBG3H1 and pAKG ricin genes, the same is not true of the RY element further downstream (Figure 16.a). Whether or not this sequence divergence causes the pCBG3H1 and pAKG ricin genes to be expressed at different levels (and/or with different developmental patterns) remains to be determined.

For the purpose of identifying any other possible cis-acting elements, the pCBG3H1 ricin gene promoter sequence was compared with sequences in the Genbank database (with the kind help of Dr MA McCrae). No further homologies were identified in this case, however.

SECTION IV.4. SIZE AND EVOLUTIONARY AFFINITIES OF THE
CASTOR BEAN LECTIN GENE FAMILY

IV.4.A. Estimated size of the lectin gene family

In order to obtain an approximate estimate of the size of the castor bean lectin gene family, a genomic Southern blot was probed with a ricin cDNA fragment and washed at low stringency. The autoradiograph in Figure 5 shows the respective size distributions of hybridising fragments in castor bean DNA digested with Eco RI and Hind III. A total of 12 bands of varying intensities could be identified in the Eco RI track but only 8 bands, also of varying intensities, could be identified in the Hind III track. This is expected, since the ricin and RCA I cDNAs both lack Hind III sites, but the RCA I cDNA contains an Eco RI site (Roberts et al, 1985). Since all the genomic clones isolated in this study, with the exception of the pCBG4H4 clone, lack a Hind III site, it is assumed that the number of bands in the Hind III lane represents fairly reliably the size of the lectin gene family. Thus the R. communis lectin gene family appears to contain approximately 8 members. The most likely reason for the varying intensities of the bands seen is that some degree of differential hybridisation is occurring, even at the low stringency used. Another possible explanation which

remains to be investigated is that one or more of the hybridising fragments might occur in multiple copies (of equal size but not necessarily nucleotide sequence) in the R. communis genome. Often, if the genome size of an organism is known, a copy number reconstruction control is included on genomic blots. In this case, however, the estimate of the R. communis genome size was not considered to be sufficiently accurate for a copy number reconstruction to be performed.

In order to assess which of the hybridising fragments might contain ricin-like genes, a second genomic blot, shown in Figure 6, was performed. The same procedure was used as with the previous Southern, except that a high stringency wash was carried out following hybridisation. It should also be noted that the Hind III and Eco RI tracks are reversed with respect to the autoradiograph in Figure 5. Two bands may be identified in each track; one of high intensity and another of lower intensity. Each corresponds to one of the more intense bands seen on the low stringency Southern. Despite this, the Eco RI fragment of approximately 1 Kb which produced an intense band on the low stringency Southern was not seen. The expected band may well have been masked, however, by a proliferation of artefactual "hot spots" which occurred further down the autoradiograph. Since no hybridising fragments smaller than 2 Kb could be identified, the area

below the 2 Kb size marker has been omitted from the photograph. Overall the Southern blot in Figure 6 suggests that there are approximately two genes in the R. communis genome which share a close homology with the ricin cDNA. As is discussed in section IV.1.C., this may account in part for the large number of potential ricin genes isolated from the partial Sau 3A library. Since the two bands seen in each track in Figure 6 are of significantly different intensities, it may be inferred that one of the ricin genes in the R. communis genome shares a greater homology with the ricin cDNA probe than the other gene. It is possible that the more weakly hybridising band seen in each case might contain a ricin E gene (section I.3.D).

From the available information, it is possible to account for 7 of the 8 proposed members of the R. communis lectin gene family. Five of the members were cloned and restriction mapped during the course of this study. One ricin gene has been fully characterised whilst the proposed second ricin gene may perhaps be present in one of the 10 lambda clones in group C. It is possible that the second ricin gene might be a ricin E variant. Alternatively there may be more than one ricin D gene as well as one (or more) ricin E gene(s). No RCA I genomic clones have been isolated, but at least one RCA I gene must be present in the R. communis genome. Thus, of the

proposed 8 members of the lectin gene family, the number of lectin genes unaccounted for may be as few as one.

IV.4.B. Linkage relationships of the castor bean lectin genes

The overall restriction maps of the DNA inserts in the lambda clones have not been determined due to time limitations. Nonetheless, at least two linkage relationships have been observed to occur. These are as follows:-

- i) The putative lectin gene partially contained in the pCBG4H4 subclone must be closely linked to the pCBG4H2 pseudogene in the R. communis genome, since both were found to occur within the DNA insert in lambda clone 7.

- ii) The pCBG3H1 (lambda clone 10) ricin gene appears to be linked with another, unidentified member of the lectin gene family. This was observed when the M13 subclone JT5 (see table 7) derived from pCBG3H1 was subjected to sequencing using an oligonucleotide primer which anneals to a region at the extreme 5' end of the ricin signal sequence. By mistake, the primer CL593, which anneals to the coding strand, was used instead of the primer CL581, which anneals at the same point but to the non-coding DNA strand (see Table 8). CL593 should not have annealed in

this case, since it was the non-coding strand which had been cloned into M13. Unexpectedly however, an AT-rich DNA sequence was seen on the autoradiograph of the sequencing gel. The only possible explanation for this observation is that pCBG3H1 contains a fragment from the 5' end of a putative lectin gene which is linked in reverse orientation to the ricin gene. The AT rich sequence seen must therefore correspond to the 5' flanking region of this putatively linked gene. Unfortunately the sequencing ladder seen was of insufficient quality to be read and no further improvement was obtained when the experiment was repeated several times. This may perhaps be due to the primer sharing only limited homology with the template used. The putatively-linked gene cannot lie on the 3' side of the full length ricin gene, since the region between the 3' end of this gene and the nearby insert/vector boundary has been fully determined (Figure 16.c) and no homologies with either of the oligonucleotides occurs in this region. The 5' end of the proposed linked gene must therefore occur at the opposite end of the pCBG3H1 DNA insert, approximately 2 Kb upstream from the ricin gene. This spacing was deduced because only a very small fragment of the second gene can be present in the pCBG3H1 and lambda clone 10 inserts, or it would have been detected earlier by Southern blotting. Approximately 2 Kb exists 5' to the full length ricin gene in the pCBG3H1 insert. The identity of the linked gene

cannot, unfortunately, be unequivocally determined by comparing the pCBG3H1 restriction map (Figure 13) with those of the other clones (Figures 11, 12 and 14). However, all the other characterised members of the lectin gene family can be ruled out as being the proposed linked gene apart from that in pCBG4H4. From the respective restriction maps of the pCBG3H1 and pCBG4H4 inserts (Figures 13 and 14), it can be seen that a Hind III site occurs both at the truncated 5' end of the putative lectin gene in pCBG4H4 and on the border of interest in the pCBG3H1 insert. This probably explains why the pCBG3H1 insert contains two separate regions which share homology with ricin signal sequence.

To conclude, it can be seen that at least 3 of the lectin gene family members are closely linked in the R. communis genome. Whether any other linkage relationships occur remains to be determined.

IV.4.C. Possible evolutionary sequence in the formation of the castor bean lectin gene family

On the basis of the results obtained in this study, a possible evolutionary sequence may be postulated for the castor bean lectin gene family. In Figure 29, this potential evolutionary sequence is depicted as a "family tree". At the foot of the "tree" is the ancestral lectin

Figure 29

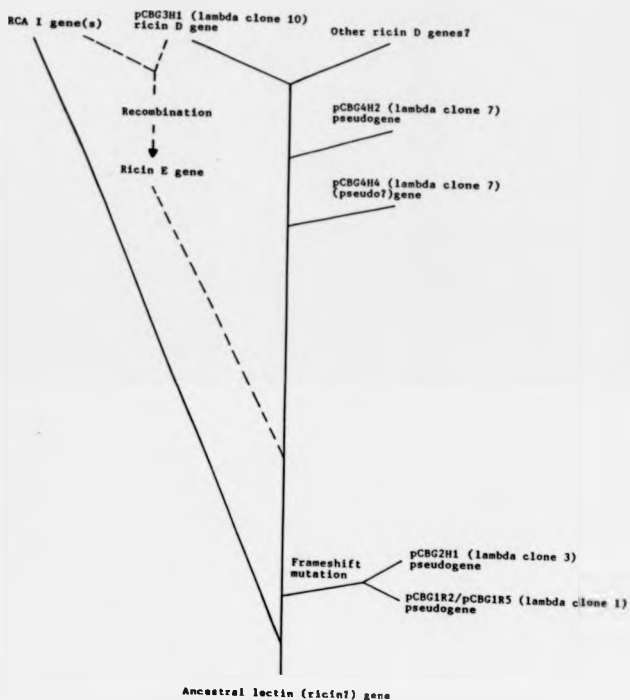


Figure 29 Possible evolutionary sequence in the
formation of the castor bean lectin gene
family

A potential evolutionary sequence, shown opposite, was postulated for the castor bean lectin gene family on the basis of the results obtained in this study. The proposed evolutionary pathway is discussed in section IV.4.C.

gene from which the other genes have presumably arisen. Ready et al (1984) have postulated that the heterodimeric lectin toxins such as ricin and abrin found in various higher plant species have evolved from the more primitive ribosome inhibiting proteins (RIPs) such as gelonin and dodecandrin also found in various higher plants. RIPs are ricin A chain-like proteins which lack a sugar binding subunit. Ready et al proposed that the genes encoding the heterodimeric lectin toxins arose during evolution by the association of an RIP gene with a gene encoding an independent sugar-binding polypeptide. It is thought that the toxin-related agglutinins (such as RCA I) found in a number of the species which produce ricin-like heterodimeric toxins have evolved from their toxin counterparts by genetic divergence. Thus it is highly likely that the ancestral gene of the R. communis lectin gene family encoded a ricin-like toxin.

Since the lambda clone 1 (pCBG1R2/pCBG1RS) and lambda clone 3 (pCBG2H1) pseudogenes show sequence affinities with both the ricin and RCA I cDNA clones, it is likely that they became non-functional at a relatively early stage in the evolution of the lectin gene family. The high degree of sequence degeneration seen in the lambda clone 1 (pCBG1R2/pCBG1RS) pseudogene sequence (Figure 17.6) adds weight to this hypothesis. The fact that the two pseudogenes share a high degree of homology and appear

to contain a common frameshift mutation (section IV.1.B) suggests that they may have diverged from a gene which was already non-functional. Thus in Figure 29 they share the same "branch" of the family tree. At the top of the tree are the functional ricin D and RCA I genes, which represent the two furthest points of sequence divergence (of the functional lectin genes) from the ancestral gene. The origin of the ricin E gene known to occur in at least some R. communis cultivars (Ladin et al., 1987) is shown by two possible pathways represented as dotted lines. The favoured theory is that the ricin E gene arose from a recombination event between a ricin D gene and an RCA I gene (Araki and Funatsu, 1987). A second possibility is that the ricin E gene has evolved in the same way as the ricin D gene, but has not diverged as far from the ancestral lectin gene. Since the lambda clone 1 pseudogene sequence shows some affinities with the RCA I cDNA sequence near the 3' terminus of the B chain region, this possibility cannot be ignored.

The pCBG4H2 (lambda clone 7) pseudogene has been placed close to the ricin D gene(s) in the tree, since it has been shown to be ricin-like, at least within the region sequenced (Figure 19). Very little is known about the putative lectin gene in pCBG4H4 (lambda clone 7), but it has been arbitrarily placed close to the pCBG4H2

pseudogene to which it is closely linked in the R.
communis genome.

SECTION IV.5. FUTURE GOALS AND FURTHER APPLICATIONS OF CLONED LECTIN GENES

Although a large part of this study was devoted to the isolation and characterisation of the various members of the castor bean lectin gene family, the long term goal of studies of this type is nevertheless to elucidate the mechanisms by which the genes of interest are regulated. The cloning, sequencing and transcriptional analysis of the functional ricin gene contained in pCBG3H1 has provided a useful basis for further experiments which will hopefully include the following:-

A. If the cis-acting DNA elements involved in the developmental and tissue-specific expression of the pCBG3H1 ricin gene are to be analysed, it is desirable that more of the promoter region be sequenced. The promoter sequence in Figure 16.a extends 250 bp upstream from the proposed cap site, but it is quite possible that some of the promoter elements of interest might occur further upstream.

B. The expression of the pCBG3H1 ricin gene in transgenic plants, under the control of its own promoter, would provide much useful information. Firstly, the identification of a ricin polypeptide in the transgenic

plants would provide unequivocal proof that the gene is functional. Secondly, the pCBG3H1 ricin gene mRNA cap and poly(A) sites could be identified far more easily in the absence of any other functional lectin genes. Thirdly it would be interesting to see whether ricin polypeptides can be safely segregated by the cells of other plant species. It is more than likely that the ricin gene promoter would retain its endosperm-specific activity in transgenic plants, but problems of toxicity could potentially make the identification of ricin polypeptides difficult. It is desirable that a plant species producing endosperm-rich seeds such as tobacco, be used as the recipient for the ricin gene, since only very low levels of lectins are produced in the cotyledons of castor bean seeds. (Harley and Beavers, 1986).

C. Ultimately, the cis-acting elements involved in the endosperm specific expression of the pCBG3H1 ricin gene must be identified by functional analysis. The most common strategy, as described in section I.4.B, is to prepare a series of promoter deletions whose activity may be measured in transgenic tissues by means of an assayable reporter gene.

D. The identification of trans-acting protein factors which bind in vivo to the pCBG3H1 ricin gene promoter would provide a first step towards elucidating the nature

of the mechanism by which the gene is "switched on and off". Studies of this type, which utilise techniques such as gel retardation (Fried and Crothers, 1981), DNase footprinting (Giles and Schmitz, 1978) or filter binding assays (Diffley and Stillman, 1986) may be used in parallel with deletion analysis to correlate the activity of specific cis-acting DNA elements with the binding of the trans-acting protein species. Although attempts are currently being made to clone the genes encoding various plant DNA binding proteins, there have to date been no reports of this being achieved. The gene encoding the mammalian $\alpha 1$ DNA binding protein has, however been cloned (Kadonaga et al, 1987) by screening a cDNA library with oligonucleotides of the same sequence as the promoter region to which the protein binds. This approach may prove useful with plant systems as it circumvents the need to purify the trans-acting factor. Other possibilities for the future include the use of transposon tagging (Motto et al, 1988) and complementation (Klee et al, 1987).

The experimental approaches described above provide an obvious continuation to the studies aimed at determining how the expression of the pCBG3H1 ricin gene is regulated. Although the central theme of these aims are addressed by such experiments, various features of the castor bean

lectin gene family nevertheless remain to be investigated. Important priorities for further investigation include:-

A. The isolation of a functional RCA I gene. Since the ricin and RCA I genes are expressed in an apparently identical fashion, it would be interesting to compare their respective promoter sequences. The isolation of an RCA I gene should be achieved by screening the amplified portion of the partial Sau 3A library at high stringency with an RCA I cDNA probe.

B. The isolation of a genomic clone containing a ricin E gene. As was discussed in section IV.4.A., there appear to be two genes in the R. communis genome which share close homology with the ricin cDNA. One of these two genes appears to be rather more weakly homologous with the ricin cDNA than the other, so it is possible that the more weakly homologous gene identified might correspond to the ricin E gene isolated by Ladin et al (1987). The identification of ricin E clones might perhaps be achieved by differential screening using appropriate fragments from the ricin and RCA I cDNAs as probes. As was discussed in section IV.4.A, it is possible that one of the lectin-positive lambda clones in group C might contain a ricin E gene.

C. Assessment of the transcriptional activity of the putative lectin gene in pCBG4H4 (lambda clone 7). A simple procedure to determine whether or not the putative lectin gene in pCBG4H4 is functional would be to carry out an RNase protection experiment using an internal fragment from the putative gene as a probe, as was done with the pCBG2H1 (lambda clone 3) and pCBG4H2 (lambda clone 7) pseudogenes.

D. Further investigation of linkage relationships. If the putative lectin gene in pCBG4H4 (lambda clone 7) is functional, then it would be desirable to isolate and sequence its promoter region. Unfortunately, the pCBG4H4 insert lacks the 5' end of the putative gene, but the region in question may be present in the pCBG3H1 plasmid (see section IV.4.B). The proposed linkage relationship could be confirmed by screening the partial Sau 3A library for clones containing fragments which span the region of interest.

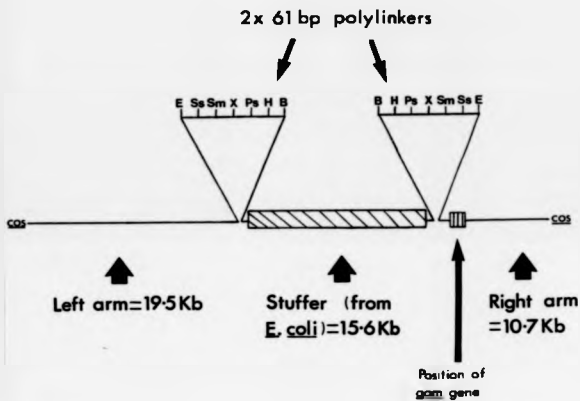
E. Isolation and characterisation of any remaining members of the lectin gene family. Even if the RCA and ricin E genes are taken into account, there must still be at least one unknown member of the lectin gene family remaining to be characterised. It is possible that lambda clone 14 (group E) might contain an as yet uncharacterised lectin gene.

SECTION IV.6. CONCLUDING REMARKS

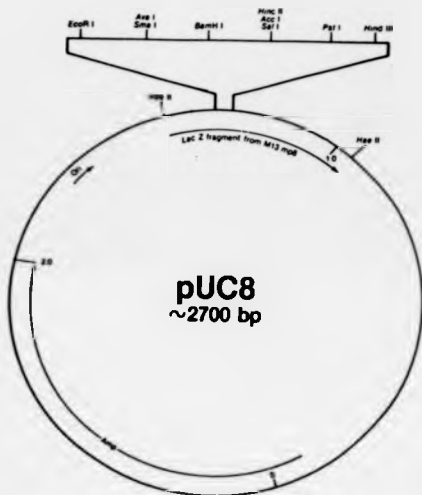
The lectin gene family of R. communis has been analysed at both a structural and a functional level. The characterisation of the nucleotide sequence and restriction site diversity of the lectin gene family, which formed a large part of this study, is a necessary prerequisite for further investigations into how the lectin genes are regulated. RNase protection analysis has enabled the identification of one functional ricin gene and three different lectin pseudogenes amongst the five putative lectin genes characterised. DNA sequencing has enabled the identification of conserved sequences potentially involved in the seed specific expression of the functional pCBG3H1 ricin gene. Priorities for future studies should include the functional elucidation of cis-acting elements in the pCBG3H ricin gene promoter and the characterisation of the other members of the lectin gene family. More broadly, it is hoped that studies of this type will increase our understanding of the structural complexity of plant genes and the processes by which their expression is regulated.

APPENDIX

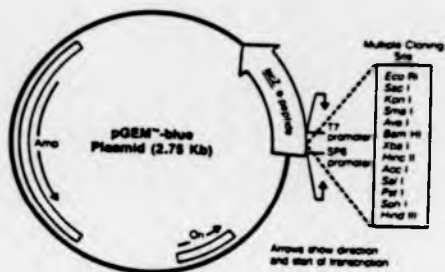
Appendix I.A. Restriction map of lambda charon 35 (Loenen and Blattner, 1983)



Appendix I.B. Restriction map of pUC8 (Vieira and
Messing, 1982)



Appendix I.C. Restriction map of pGEM-blue plasmids
(Promega Biotech)



REFERENCES

- Altachul A.M., Yatsu L.Y., Ory R.L. and Engleman E.M. (1966). Seed proteins. *Ann. Rev. Plant Physiol.* 17: 113-136.
- An G. (1987). A potential Z-DNA forming sequence is an essential upstream element of a plant promoter. *Bioassays* 7: 211-214.
- Araki T. and Funatsu G. (1987). The complete amino acid sequence of the B chain of ricin E isolated from small grain castor bean seeds. Ricin E is a gene recombination product of ricin D and Ricinus communis agglutinin. *Biochim. Biophys. Acta* 911: 191-200
- Avery O.T., Macleod C.M. and McCarty M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from Pneumococcus Type III. *J. Exp. Med* 79: 137-158.
- Baumann G., Raschke E., Bevan M. and Schoffl F. (1987) Functional analysis of sequences required for transcriptional activation of a soybean heat shock gene in transgenic tobacco plants. *EMBO J.* 6: 1161-1166.
- Baumlein H., Wobus U., Pustell J. and Kafatos F.C. (1986). The legumin gene family: structure of a B type gene of Vicia faba and a possible legumin gene specific regulatory element. *Nucl. Acids Res.* 14: 2707-2720.
- Beachy R.N., Chen Z.-L., Horsch R.B., Rogers S.G.,

- Hoffmann N.J. and Fraley R.T. (1985). Accumulation and assembly of soybean B-conglycinin in seeds of transformed petunia plants. *EMBO J.* 4: 3047-3053.
- Bender W., Spierer P. and Hogness D.S. (1983). Chromosomal walking and jumping to isolated DNA from the Ace and roxy loci and the bithorax complex in Drosophila melanogaster. *J. Mol. Biol.* 168: 17-34.
- Benoist C. and Chambon P. (1981). In vivo sequence requirements of the SV40 early promoter region. *Nature* 290: 304-310.
- Benoist C., O'Harek., Breathnach R. and Chambon P. (1980). The ovalbumin gene-sequence of putative control regions. *Nucl. Acids Res.* 8: 127-142
- Benton W.D. and Davis R.W. (1977). Screening lambda gt recombinant clones by hybridization to single plaques in situ. *Science* 196: 180-182.
- Berget S.M. (1984). Are U4 small nuclear ribonucleoproteins involved in polyadenylation? *Nature* 309: 179-182.
- Bevan M. (1984). Binary Agrobacterium vectors for plant transformation. *Nucl. Acids Res.* 12: 8710-8720.
- Biggin M.D., Gibson T.J. and Hong G.F. (1983). Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* 80: 3963-3965.
- Birnboim H.C. and Doly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid

- DNA. Nucl. Acids Res. 7: 1513.
- Blattner F.R., Williams B.G., Blechl A.E., Denniston-Thompson K., Faber H.E., Furlong L.-A., Grunwald D.J., Kiefer D.O., Moore D.D., Schumm J.W., Sheldon E.L. and Smithies O. (1977). Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science 196: 161-169.
- Bolivar F., Rodriguez R.L., Greene P.J., Betlach M.C., Heynecker H.L., Boyer H.W., Cross J.H. and Falkow S. (1977). Construction and characterization of new cloning vehicles II. A multipurpose cloning system. Gene 2: 95.
- Brogliè R., Coruzzi G., Lamma G., Keith B. and Chua N.-H. (1983). Structural analysis of nuclear genes coding for the precursor to the small subunit of wheat ribulose 1, 5-bisphosphate carboxylase. Biotechnology 1: 55-61.
- Brown J.W.S., Feix G. and Frendewey D. (1986). Accurate in vitro splicing of two pre-mRNA plant introns in a HeLa cell nuclear extract. EMBO J. 5: 2749-2758
- Bull H., Li S.-L., Fowler E. and Lin T.-T.S. (1980). Isolation and characterisation of cyanogen bromide fragments of the A and B chains of the anti-tumor toxin ricin D. Int. J. Peptide Protein Res. 16: 208-218.
- Cawley D.B., Hedblom M.L. and Houston L.L. (1978). Homology between ricin and Ricinus communis agglutinin: amino terminal sequence analysis and

- protein synthesis inhibition studies. Arch. Biochem. Biophys. 190: 744-755.
- Chen Z.-L., Pan N.-S. and Beachy R.N. (1988). A DNA sequence that confers seed-specific enhancement to a constitutive promoter. EMBO J. 7: 297-302
- Chen Z.L., Schuler M.A. and Beachy R.N. (1986). Functional analysis of regulatory elements in a plant embryo-specific gene. Proc. Natl. Acad. Sci. USA 83: 8560-8564.
- Colot V., Roberts L.S., Kavanagh T.A., Bevan M.W. and Thompson R.D. (1987). Localisation of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO J. 6: 3559-3564.
- Dazzo F.B. and Sherwood J.E. (1983). Trifoliin A: A Rhizobium recognition lectin in white clover roots. In: Chemical Taxonomy, Molecular Biology and function of Plant Lectins (Eds. I.J. Goldstein and M.E. Etxler) pp. 209-223 Alan R. Liss, New York.
- Dean C., Tamaki S., Dunsmuir P., Faureau M., Katayama C., Dooner H. and Bedbrook J. (1986). mRNA transcripts of several plant genes are polyadenylated at multiple sites in vivo. Nucl. Acids Res. 14: 2229-2240
- Deikman J. and Fischer R.L. (1988). Interaction of a DNA binding factor with the 5'-flanking region of an ethylene- responsive fruit ripening gene from tomato. EMBO J. 7: 3315-3320.
- De La Pena A., Lorz H. and Schell J. (1987). Transgenic

- rye plants obtained by injecting DNA into young floral tillers. *Nature* 325: 274-276.
- Diaz C. L., Malchers L.S., Hooykas P.J.J., Lugtenberg B.J.J. and Kijne J.W. (1989). Root lectin as a determinant of host-plant specificity in the Rhizobium-legume symbiosis. *Nature* 338: 579-581
- Dickinson C.D., Evans R.P. and Nielson M.C. (1988). RY repeats are conserved in the 5'-flanking regions of legume seed-protein genes. *Nucl. Acids Res.* 16: 371
- Diffley J.F.X. and Stillman B. (1986). Purification of a cellular, double-stranded DNA binding protein required for the initiation of Adenovirus DNA replication by using a rapid filter-binding assay. *Mol. Cell. Biol.* 6: 1363-1373.
- Doyle J.J., Schuler M.A., Godette D., Zenger V., Beachy R.N. and Slightom J.L. (1986). The glycosylated seed storage proteins of Glycine max and Phaseolus vulgaris. Structural homologies of genes and proteins. *J. Biol. Chem.* 261: 9228-9238
- Drouin G. and Dover G.A. (1987). A plant processed pseudogene. *Nature* 328: 557-558.
- Drouin J. (1980). Cloning of human mitochondrial DNA in Escherichia coli. *J. Mol. Biol.* 140: 15
- Dugaiczyk A., Boyer H.W. and Goodman H.M. (1975). Ligation of Eco RI endonuclease-generated DNA fragments into linear and circular structures. *J. Mol. Biol.* 96: 171-184.
- Dynan W.S. and Tjian R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-

- binding proteins. *Nature* 316: 774-778.
- Efstathiou A., Kafatos F.C., Maxam A.M. and Maniatis T. (1976). Enzymatic in vitro synthesis of globin genes. *Cell* 7: 279-288.
- Ellis J.R., Shirsat A.H., Hopher A., Yarwood J.N., Gatehouse J.A., Croy R.R.D. and Boulter D. (1988). Tissue-specific expression of a pea legumin gene in seeds of Nicotiana plumbaginifolia. *Plant Mol. Biol.* 10: 203-214.
- Everett R.D., Batty D. and Chambon P. (1983). The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. *Nucl. Acids Res.* 11: 2447-2464.
- Fabijanski S. and Altosaar I. (1985). Evidence for translational control of storage protein biosynthesis during embryogenesis of Avena sativa L. (oat) endosperm. *Plant Mol. Biol.* 4: 211-218
- Feinberg A.P. and Vogelstein B. (1983). A technique of radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132: 6-13.
- Fitzgerald M. and Shenk T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 24: 251-260.
- Flores S. and Tobin E.M. (1988). Cytokinin modulation of LHCP mRNA levels: the involvement of post-transcriptional regulation. *Plant Mol. Biol.* 11: 409-415

- Fluhr R., Kuhlemeier C., Nagy F. and Chua N.-H. (1986). Organ- specific and light-induced expression of plant genes. *Science* 232: 1106-1112.
- Forde B.G., Hayworth A., Pywell J. and Kreis M. (1985). Nucleotide sequence of a B1 hordein gene and the identification of possible regulatory elements in endosperm storage protein genes from barley, wheat and maize. *Nucl. Acids Res.* 13: 7327-7339.
- Fried M. and Crothers D.M. (1982). Equilibria and kinetics of lac repressor interactions by polyacrylamide gel electrophoresis. *Nucl. Acids Res.* 9: 6505-6525.
- Fromm M.E., Taylor L.P. and Walbot V. (1985). Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proc. Natl. Acad. Sci. USA* 82: 5824-5828.
- Funatsu G., Kimura M. and Funatsu M. (1979). Primary structure of the Ala chain of ricin D. *Agric. Biol. Chem.* 43: 2221-2224.
- Funatsu G., Yoshitakes S. and Funatsu M. (1978). Primary structure of the Ile chain of ricin D. *Agric. Biol. Chem.* 42: 501-503.
- Galas D.J. and Schmitz A. (1978). DNase footprinting: a for the detection of protein-DNA binding specificity. *Nucl. Acids Res.* 5: 3157-3170.
- Gallagher T.F. and Ellis R.J. (1982). Light-stimulated transcription of genes for two chloroplast polypeptides

- in isolated pea leaf nuclei. EMBO J. 1: 1493-1498.
- Gallie D.R., Sleat D.E., Watts J.W., Turner P.C. and Wilson T.M.A. (1987). The 5'-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts in vitro and in vivo. Nucl. Acids Res. 15: 3257-3273
- Gifford D.J., Greenwood J.S. and Bawley J.D. (1982). Deposition of matrix and crystalloid storage proteins during protein body development in the endosperm of Ricinus communis L. cv. Hale seeds. Plant Physiol. 69: 1471-1478
- Goldberg R.B., Hoschek G. and Vodkin L.O. (1983). An insertion sequence blocks the expression of a soybean lectin gene. Cell 33: 456-475.
- Goldstein I.J. and Hayes C.E. (1978). The lectins: carbohydrate binding proteins of plants and animals. Advances in Carbohydrate Chemistry and Biochemistry 35: 127-340.
- Gonatas J., Stieber A., Olsnes S. and Gonatas N.K. (1980). Pathways involved in fluid phase and adsorptive endocytosis in neuroblastoma. J. Cell Biol 87: 579-588.
- Gonatas N.K., Stieber A., Kim S.V., Graham D.I. and Avrameas S. (1975). Internatization of neuronal plasma membrane ricin receptors into the Golgi apparatus. Expl. Cell Res. 94: 426-431.
- Green P.J., Kay S.A. and Chua N.-H. (1987). Sequence-

- specific interactions of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene. *EMBO J.* 6: 2543-2549.
- Grierson D. (1982). RNA processing and other post-transcriptional modifications. In: *Nucleic Acids and Proteins in Plants II. Structure, Biochemistry and Physiology of Nucleic Acids* (eds. B. Parthier and D. Boulter): pp 192-223. Springer-Verlag, Berlin.
- Grosschedl R. and Birnstiel M.L. (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. *Proc. Natl. Acad. Sci. USA* 77: 1432-1436.
- Grosveld G.C., De Boer E., Shewmaker C.K. and Flavell R.A. (1982). DNA sequences necessary for transcription of the rabbit beta-globin gene in vivo. *Nature* 295: 120-126.
- Ha S.-B. and An G. (1988). Identification of upstream regulatory elements involved in the development expression of the Arabidopsis thaliana cab 1 gene. *Proc. Natl. Acad. Sci. USA* 85: 8017-8021.
- Halling K.C., Halling A.C., Murray E.E., Ladin B.F., Houston L.L. and Weaver R.F. (1985). Genomic cloning and characterisation of a ricin gene from Ricinus communis. *Nucl. Acids. Res.* 13: 8019-8033
- Hanahan D. (1983). Studies in transformation of Escherichia coli with plasmids. *J. Mol. Biol* 166: 557-

- 580.
- Hanley B.A. and Schuler M.A. (1988). Plant intron sequences: evidence for distinct groups of introns. *Nucl. Acids Res.* 16: 7159-7176.
- Harley S.M. and Beevers H. (1986). Lectins in castor bean seedlings. *Plant Physiol.* 80: 1-6
- Hashimoto C. and Steitz J.A. (1986). A small nuclear ribonucleoprotein associates with the AAUAAA polyadenylation signal in vitro. *Cell* 45: 581-591.
- Hattori T. and Nakamura K. (1988). Genes coding for the major tuberous protein of sweet potato: identification of putative regulatory sequences in the 5' upstream region. *Plant Mol. Biol.* 11:417-426
- Herrmann A., Schultz W. and Hahlbrock K. (1988). Two alleles of the single-copy chalcone synthase gene in parsley differ by a transposon-like element. *Mol. Gen. Genet.* 212: 93-98
- Higgins T.J.V., Chandler P.M., Zurawski G., Button S.C. and Spencer D. (1982). The biosynthesis and primary structure of pea seed lectin. *J. Biol. Chem.* 258: 9544-9549.
- Hoffman L.M. (1984). Structure of a chromosomal Phaseolus vulgaris lectin gene and its transcript. *J. Mol. Appl. Genet.* 2: 447-453.
- Hoffman L.M. and Donaldson D.D. (1985). Characterization of two Phaseolus vulgaris phytohaemagglutinin genes closely linked on the chromosome. *EMBO J.* 4: 883-889

- Hoffman L.M., Donaldson D.D., Bookland R., Rashka K. and Herman E.M. (1987). Synthesis and protein body deposition of maize 15-Kd zein in transgenic tobacco seeds. *EMBO J.* 6: 3213-3221.
- Houston L.L. and Dooley T.P. (1982). Binding of two molecules of 4-methylumbelliferyl galactose or 4-methylumbelliferyl N-acetylgalactosamine to the B chains of ricin and *Ricinus communis* agglutinin and to purified ricin B chain. *J. Biol. Chem.* 257: 4147-4151.
- Hu N.-T. and Messing J. (1982). The making of strand-specific probes. *Gene* 17: 271-277
- Hu N.-T., Peifer M.A., Heidecker G., Messing J. and Rubenstein I. (1982). Primary structure of a genomic zein sequence of maize. *EMBO J.* 1: 1337-1342.
- Hughes D.W. and Galau G.A. (1987). Translation efficiency of *Lea* mRNAs in cotton embryos: minor changes during embryogenesis and germination. *Plant Mol. Biol.* 9: 301-313
- Huynh T.V., Young R.A. and Davis R.W. (1985). Construction and screening of genomic libraries in lambda gt10 and lambda gt11. In: *DNA Cloning Volume II* (ed. D.M. Glover), IRL Press.
- Jefferson R.A., Kavanagh T.A. and Bevan M.W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6: 3901-3907.
- Jenson E.O., Marcker K.A., Schell J. and De Bruijn F.J.

- (1988). Interaction of a nodule specific, trans acting factor with distinct DNA elements in the soybean leghaemoglobin lbc3 5' upstream region. *EMBO J.* 7: 1265-1271.
- Jofuku K.D., Okamuro J.K. and Goldberg R.B. (1987). Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. *Nature* 328: 734-737.
- Joshi C.P. (1987). Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. *Nucl. Acids Res.* 15: 9627-9640.
- Kaminski P.A., Buffard D. and Strosberg A.D. (1987). The pea lectin gene family contains only one functional gene. *Plant Mol. Biol.* 9: 497-507.
- Kao, F.-T., Hartz J.A., Law M.L. and Davidson J.N. (1982). Isolation and chromosomal localization of unique DNA sequences from a human genomic library. *Proc. Natl. Acad. Sci. USA* 79: 865-869.
- Karn J., Brenner S., Barnett L. and Cesareni G. (1980). Novel bacteriophage cloning vector. *Proc. Natl. Acad. Sci. USA* 77: 5172-5176.
- Keith B. and Chua N.-H. (1986). Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco. *EMBO J.* 5: 2419-2425.
- Keller G.H. and Taylor J.M. (1976). Effect of hypophysectomy on the synthesis of rat liver albumin. *J. Biol. Chem.* 251: 3768-3773.

- Khorana H.G. (1968). Nucleic acid synthesis in the study of the genetic code. In: Nobel Lectures: Physiology or Medicine (1963-1970) pp. 372-395, American Elsevier (1973).
- Khoury G. and Gruss P. (1983). Enhancer elements. *Cell* 33: 313-314.
- Kislev N. and Rubenstein I. (1980). Utility of ethidium bromide in the extraction from whole plants of high molecular weight maize DNA. *Plant Physiol.* 66: 1140-1143.
- Klee H.J., Hayford M.B. and Rogers S.G. (1987). Gene rescue in plants: A model system for "shotgun" cloning by retransformation. *Mol. Gen. Genet.* 210: 282-287
- Klein T.M., Gradziel T., Fromm M.E. and Sanford J.C. (1988). Factors influencing gene delivery into Zea mays cells by high-velocity microprojectiles. *Bio/technology* 6: 559-563.
- Kozak M. (1980). Role of ATP in binding and migration of 40S ribosomal subunits. *Cell* 22: 459-467.
- Krebbers K., Herdies L., De Clercq A., Seurinck J., Leemans J., Van Montagu M. and Vandekerckhove J. (1988) Determination of the processing sites of an Arabidopsis 2S albumin and characterisation of the complete gene family. *Plant Physiol.* 87: 859-866
- Kreig P.A. and Melton D.A. (1987). In vitro RNA synthesis with SP6 RNA polymerase. *Methods Enzymol.* 155: 397-415.

- Krens F.A., Molendijk L., Willems G.J. and Schilpercoort R.A. (1982). In vitro transformation of plant protoplasts with Ti-plasmid DNA. *Nature* **296**: 72-74.
- Ladin B.F., Murray E.E., Halling A.C., Halling K.C., Tilakaratne N., Long G.L., Houston L.L. and Weaver R.F. (1987). Characterisation of a cDNA encoding ricin E, a hybrid ricin-Ricinus communis agglutinin gene from the castor oil plant Ricinus communis. *Plant Mol. Biol.* **9**: 287-295.
- Lamb F.I. (1984). PhD Thesis, University of Warwick.
- Lamb F.I., Roberts L.M. and Lord J.M. (1985). Nucleotide sequence of cDNA coding for preproricin. *Eur. J. Biochem.* **148**: 265-270.
- Liener I.E. (1976). Phytoagglutinins (phytolectins). *Ann. Rev. Plant Physiol.* **27**: 291-319.
- Lindsay K. and Jones M.G.K. (1987). Transient gene expression in electroporated protoplasts and intact cells of sugar beet. *Plant Mol. Biol.* **10**: 43-52.
- Lipphardt S., Brettschneider R., Kreuzaler F., Schell J. and Dangl J.L. (1988). UV-inducible transient expression in parsley protoplasts identifies regulatory cis-elements of a chimeric Antirrhinum majus chalcone synthase gene. *EMBO J.* **7**: 4027-4033.
- Loenen W.A.M. and Blattner F.R. (1983). Lambda charon vectors (Ch32, 33, 34 and 35) adapted for DNA cloning in recombination-deficient hosts. *Gene* **26**: 171-179.
- Lord J.M. (1985). Precursors of ricin and Ricinus

- communis agglutinin. Glycosylation and processing during synthesis and intracellular transport. Eur. J. Biochem. 146: 411-416
- Lord J.M., Roberts L.M., Thorpe P.E. and Vitetta E.S. (1985). Immunotoxins. Trends Biotechnol. 3: 175-179
- Lorz H., Baker B. and Schell J. (1985). Gene transfer to cereal cells mediated by protoplast transformation. Mol. Gen. Genet. 199: 178-182.
- Lycett G.W., Croy R.R.D., Shirsat A.H., Richards D.M. and Boulter D. (1985). The 5'-flanking regions of three pea legumin genes: comparison of the DNA sequences. Nucl. Acids Res. 13: 6733-6743.
- Maier U.-G., Brown J.W.S., Toloczky C. and Feix G. (1987). Binding of a nuclear factor to a consensus sequence in the 5' flanking region of zein genes from maize. EMBO J. 6: 17-22.
- Mandel M. and Higa A. (1970). Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. 53: 159-162.
- Maniatis T., Fritsch E.F. and Sambrook J. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York, USA.
- Maniatis T., Hardison R. C., Lacy E., Lauer J., O'Connell C. and Quon D. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell 15: 687-701
- Maxam A.M. and Gilbert W. (1977). A new method for

- sequencing DNA. Proc. Natl. Acad. Sci. USA 74: 560-564
- McLauchlan J., Gaffney D., Whitton J.L. and Clements J.B. (1985). The consensus sequence YGTGTTY located downstream the AATAAA signal is required for efficient formation of mRNA 3' termini. Nucl. Acids Res. 13: 1347-1368.
- Messing J., Geraghty D., Heidecker G., Hu N., Kridl J. and Rubenstein, I. (1983). Plant gene structure In: Genetic Engineering of Plants (Kosuge, T., Meredith, C.P. and Hollaender A., eds.), pp. 211-227, Plenum Publishing Corp., New York.
- Messing J. and Vieira J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19: 269-276.
- Mirelman D.E., Galun E., Sharon N. and Lotan R. (1975). Inhibition of fungal growth by wheat germ agglutinin. Nature 256 416-418.
- Mise T., Funatsu G., Ishiguro M. and Funatsu M. (1977). Isolation and characterization of ricin E from castor beans. Agric. Biol. Chem. 41: 2041-2046.
- Morelli G., Nagy F., Fraley R.T., Rogers S.G. and Chua N.-H. (1985). A short conserved sequence is involved in the light-inducibility of a gene encoding ribulose 1,5-bisphosphate carboxylase small subunit of pea. Nature 315: 200-204.
- Murray M.G., Kennard W.C., Drong R.F. and Slightom J.L. (1984). Use of a recombination-deficient phage lambda

- system to construct wheat genomic libraries. *Gene* 30: 237-240.
- Negoro T., Momma T. and Fukazawa C. (1985). A cDNA clone encoding a glycinin subunit precursor of soybean. *Nucl. Acids Res.* 13: 1341-1350
- Nicolson G.L., Blaustein J. and Etzler M.Ek. (1974). Characterization of two plant lectins from Ricinus communis and their interaction with a murine lymphoma. *Biochem.* 13: 196-204.
- Odell J.T., Knowlton S., Lin W. and Mauvais J. (1988). Properties of an isolated transcription stimulating sequence derived from the cauliflower mosaic virus 35S promoter. *Plant Mol. Biol.* 10: 263-272
- Odell J.T., Nagy F. and Chua N.-H. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* 313: 810-812
- Okayama H. and Berg P. (1982). High-efficiency cloning of full-length cDNA. *Mol. Cell. Biol.* 2: 161-170.
- Olanes S. and Pihl A. (1976). Abrin, ricin and their associated agglutinins. In: *The Specificity of Action of Animal, Bacterial and Plant Toxins* (ed. P. Cuatrecasas) pp. 129-173, Chapman and Hall.
- Olanes S. and Pihl A. (1982). Toxic lectins and related proteins. In: *Molecular Actions of Toxins and Viruses* (eds. P. Cohen and S. Van Heyningen) pp. 51-105, Elsevier Biomedical Press.

- Olsnes S., Refsum K. and Pihl A. (1974A). Mechanism of action of the toxic lectins abrin and ricin. *Nature* 249: 627-631.
- Olsnes S., Saltvedt E. and Pihl A. (1974B). Isolation and comparison of galactose binding lectins from Abrus precatorius and Ricinus communis. *J. Biol. Chem.* 249: 803-810.
- Olsnes S., Sandvig K. and Pihl A. (1979). In: *Protides of the Biological Fluids, 27th Symposium* (ed. H. Peters) pp. 395-399, Pergamon Press.
- O'Neal J.K., Pokalsky A.R., Kishne K.L. and Shawmaker C.K. (1987). Isolation of tobacco SSU genes: characterization of a transcriptionally active pseudogene. *Nucl. Acids Res* 15: 8661-8677.
- Pappenheimer A.M., Olsnes S. and Harper A.A. (1974). Lectins from Abrus precatorius and Ricinus communis I. Immunochemical relations between toxins and agglutinins. *J. Immunol.* 113: 835-841.
- Paskowski J., Shillito R.D., Saul M., Mandak V., Hohn T., Hohn B. and Potrykus I. (1984). Direct gene transfer to plants. *EMBO J.* 3: 2717-2722.
- Podder S.K., Surolia A. and Bachhawat B.K. (1974). On the specificity of carbohydrate-lectin recognition. The interaction of a lectin from Ricinus communis beans with simple saccharides and concanavalin A. *Eur. J. Biochem.* 44: 151-160.
- Proudfoot N.J. (1979). Eukaryotic promoters? *Nature* 279:

376.

Proudfoot N.J. (1984). The end of the message and beyond.
Nature 307: 412-413.

Refnes K., Olanes S. and Pihl A. (1974). On the toxic
proteins abrin and ricin. Studies on their binding to
and entry into Ehrlich ascites cells. J. Biol. Chem.
249: 3557-3562.

Rigby P.W.J., Dieckmann M., Rhodes C. and Berg P. (1977).
Labelling deoxyribonucleic acid to high specific
activity in vitro by nick translation with DNA
Polymerase I. J. Mol. Biol. 113 237-251.

Roberts L.M., Lamb F.I., Pappin D.J.C. and Lord J.M.
(1985). The primary sequence of Ricinus communis
agglutinin. Comparison with ricin. J. Biol. Chem.
260: 15683-15686

Roberts L.M. and Lord J.M. (1979A.). Ribonucleic acid
synthesis in germinating castor bean endosperm. J.
Exp. Bot. 30: 739-749

Roberts L.M. and Lord J.M. (1979B.). Developmental
changes in the activity of messenger RNA isolated from
germinating castor bean endosperm. Plant Physiol. 64:
630-634

Roberts L.M. and Lord J.M. (1981A.). Protein biosynthetic
capacity in the endosperm tissue of ripening castor
bean seeds. Planta 152: 420-427

Roberts L.M. and Lord J.M. (1981B.). The synthesis of
Ricinus communis agglutinin. Cotranslational and

- postranslational modification of agglutinin polypeptides. *Eur. J. Biochem.* 119: 31-41
- Roussell D.L., Boston R.S., Goldsbrough P.B. and Larkins B.A. (1988). Deletion of DNA sequences flanking an Mr 19 000 zein gene reduces its transcriptional activity in heterologous plant tissues. *Mol. Gen. Genet.* 211: 202-209.
- Roychoudhury R., Jay E. and Wu R. (1976). Terminal labelling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. *Nucl. Acids Res.* 3: 863-877.
- Sanchez-Serrano J.J., Keil M., O'Connor A., Schell J. and Willmitzer L. (1987). Wound-induced expression of a potato proteinase inhibitor II gene in transgenic tobacco plants. *EMBO J.* 6: 303-306.
- Sanduig K., Olsson S. and Pihl A. (1978). Binding, uptake and degradation of the toxic proteins abrin and ricin by toxin-resistant cell variants. *Eur. J. Biochem.* 82: 13-23.
- Sanger F., Nicklen S. and Coulson A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463-5467.
- Schachter H. (1974). The subcellular sites of glycosylation. *Biochem. Soc. Symp.* 40: 57-71.
- Scherthaner J.P., Matzke M.A. and Matzke A.J.M. (1988). Endosperm-specific activity of a zein gene promoter in transgenic tobacco plants. *EMBO J.* 7: 1249-1255.

- Schreier P.H., Seftor E.A., Schell J. and Bohnert H.J. (1985). The use of nuclear-encoded sequences to direct the light-regulated synthesis and transport of a foreign protein into plant chloroplasts. *EMBO J.* 4: 25-32.
- Seed B., Parker R.C. and Davidson N. (1982). Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. *Gene* 19: 201-209.
- Simpson J., Schell J., Van Montagu M. and Herrera-Estrella L. (1986). Light-inducible and tissue-specific pea lhcp gene expression involves an upstream element combining enhancer- and silencer-like properties. *Nature* 323: 551-554.
- Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98 503-517.
- Stillmark H. (1889). *Arbeiten Pharmakologischen Institutes Zu Dorpat* 3: 59.
- Traina V. and Cohen J.G. (1983). The dot blot assay for RNA or how the blot got its spots. *Focus* 5 (1): 12
- Ueng P., Galili G., Sapanara V., Goldsbrough P.B., Dube P., Beachy R.N. and Larkins B.A. (1988). Expression of a maize storage protein gene in *Petunia* plants is not restricted to seeds. *Plant Physiol.* 86: 1281-1285
- Van den Broeck G., Timko M.P., Kausch A.P., Cashmore A.R., Van Montagu M. and Herrera-Estrella L. (1985).

Targetting of a foreign protein to chloroplasts by fusion to the transit peptide from the small subunit of ribulose 1,5-bisphosphate carboxylase. *Nature* 313: 358-363

Vanin E.F. (1984). Processed pseudogenes.

Characteristics and evolution. *Biochim. Biophys. Acta* 782: 231-241

Vanin E.F., Goldberg G.I., Tucker P.W. and Smithies O.

(1980). A mouse alpha-globin-related pseudogene lacking intervening sequences. *Nature* 286: 222-226.

Van Santen V.L. and Spritz R.A. (1987). Splicing of plant pre-mRNAs in animal systems and vice versa. *Gene* 56: 253-265.

Vesterberg O. (1975). In: Isoelectric focusing (eds. J.P. Arbuthnott and J.A. Beeley), pp. 78-96, Butterworths, London.

Vieira J. and Messing J. (1982). The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19: 259-268.

Villafranca J.E. and Robertus J.D. (1981). Ricin B chain is a product of gene duplication. *J. Biol. Chem.* 256: 554-556.

Viotti A., Sala E., Marotto R., Alberi P., Balducci C. and Soave C. (1979). Genes and mRNAs coding for zein polypeptides in *Zea mays*. *Eur. J. Biochem.* 102: 211-222.

- Vodkin L.O., Rhodes P.R. and Goldberg R.B. (1983). cA lectin insertion has the structural features of a transposable element. *Cell* 34: 1023-1031.
- Voelker T.A., Staswick P. and Chrispeels M.J. (1986). Molecular analysis of two phytohaemagglutinin genes and their expression in Phaseolus vulgaris cv. Pinto, a lectin- deficient cultivar of the bean. *EMBO J.* 5: 3075-3082.
- Von Heijne G. (1983). Patterns of amino acids near signal sequence cleavage sites. *Eur. J. Biochem.* 133: 17-21.
- Wahl G.M., Stern M. and Stark G.R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzylxymethyl-paper and rapid hybridization by using dextran sulfate. *Proc. Natl. Acad. Sci. USA* 76: 3683-3687.
- Wang Y.-C., Klein T.M., Fromm M., Cao J., Sanford J.C. and Wu R. (1988). Transient expression of foreign genes in rice, wheat and soybean cells following particle bombardment. *Plant Mol. Biol.* 11: 433-439.
- Weaver R.F. and Weissmann C. (1979). Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S beta-globin mRNA precursor and mature 10S beta-globin mRNA have identical map coordinates. *Nucl. Acids. Res.* 7: 1175-1193.
- Werr W., Frommer W.-B., Maas C. and Starlinger P. (1985). Structure of the sucrose synthase gene on chromosome 9 of Zea mays L. *EMBO J.* 4: 1373-1380.

- Wickens M. and Stephenson P. (1984). Role of the conserved AAUAAA sequences: four AAUAAA point mutants prevent messenger RNA 3' end formation. *Science* 226: 1045-1051.
- Williams J.G. and Mason P.J. (1985). Hybridisation in the analysis of RNA. In: DNA cloning volume II (ed. D.M. Glover), IRL Press.
- Wood W.B. (1966). Host specificity of DNA produced by Escherichia coli: bacterial mutations affecting the restriction and modification of DNA. *J. Mol. Biol.* 16: 118-133.
- Young R.A. and Davis R.W. (1983). Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. USA* 80: 1194-1198.
- Zambryski P., Holsters M., Kruger K., Depicker A., Schell J., Van Montagu M. and Goodman H. (1980). Tumor DNA structure in plant cells transformed by A. tumefaciens. *Science* 209: 1385-1391.

THE BRITISH LIBRARY DOCUMENT SUPPLY CENTRE

TITLE

THE LECTIN GENE FAMILY OF RICINUS COMMUNIS

AUTHOR

James W. Tregear,

INSTITUTION
and DATE

University of Warwick

1989

Attention is drawn to the fact that the copyright of this thesis rests with its author.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no information derived from it may be published without the author's prior written consent.



THE BRITISH LIBRARY
DOCUMENT SUPPLY CENTRE
Boston Spa, Wetherby
West Yorkshire
United Kingdom

20

REDUCTION X

CAMERA

2

D90795