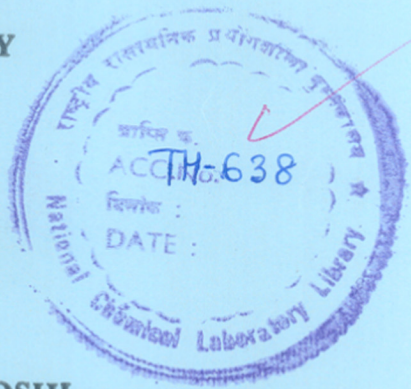
CLONING AND ANALYSIS OF GLUTELIN: A MAJOR SEED STORAGE PROTEIN IN BASMATI RICE

A THESIS
SUBMITTED TO THE
UNIVERSITY OF POONA
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
(IN CHEMISTRY)



BY
MRS. KALPANA SHRIKANT JOSHI
M. Sc.

DIVISION OF BIOCHEMICAL SCIENCES
NATIONAL CHEMICAL LABORATORY
PUNE-411 008 (INDIA)

1991

dedicated to my family

CONTENTS

CHAPTER 3 : RESULTS

SECTION I

3.1 Construction of rice (Basraati 370) genomic library in Charon40 vector

SECTION II

3.2 Screening of the rice genomic library and identification of clones containing glutelin gene

SECTION III

3.3 Characterization of rice (Basmati-370) glutelin gene

ACKNOWLEDGEMENT

I consider myself extremely fortunate to have Dr. P.K.Ranjekar as my research guide. It is with all sincerity that I express my indebtedness to him for introducing me to the field of Plant Molecular Biology and for an inspiring guidance, moral support and encouraging advice.

I wish to express my sincere thanks to Dr.(Mrs.) Vidya Gupta for taking an active interest in the present work. Her kind advice and generous help in solving all possible problems I encountered during my Ph.D work have contributed immensely towards my progress and success.

I do not wish to be formal, but I feel happy to record my special thanks to Vivek for his friendly help during the tenure of this work.

I am deeply indebted to Dr.(Mrs) Meena Lagu, Dr.(Mrs) Shubhada Tamhankar and Mrs. Madhavi Dabak for their personal help and friendly guidance.

The help rendered by Mrs.M.N.Sainani, Vaijayanti, Mr. P.K.Chitnis, Mr.B.G.Patil and Dr. Yogesh Mawal was extremely useful and I duly acknowledge them for their cooperation.

I will always remember the jolly companionship of my senior colleagues Dr.(Mrs) N.A.Sahasrabudhe, Dr. (Mrs) Lalita, Dr. R.R.Hendre and Dr.(Miss) Radha Parkhi and friends Gagan, Suman, Madhur, Ramakrishna, Sujata, Renuka, Urmila, Savita, Kamala and Vishwas.

I gratefully acknowledge the technical assistance

given by Mr.S.T.Dhume, and Mr.M.N.Kamathe in maintaining the various instruments.

It gives me great pleasure to thank Mr. Ajit Ranjekar for his enthusiastic typing of this thesis. I am also thankful to Miss Usha Deshpande for typing the bibliography.

I am also thankful to Dr. John Barnabas for his encouragement and help in his capacity as Head, Division of Biochemical Sciences and the Director, N.C.L, for his kind permission to submit this work in the form of a thesis.

The financial help in the form of a Research Fellowship from the Council of Scientific and Industrial Research (CSIR), New Delhi is gratefully acknowledged.

Last, but not the least, I owe to the fullest extent to my parents and sister Minti for their inspiration and support and also to my parents-in-law for their full hearted cooperation. A special appreciation is due to my husband, Dr. Shrikant and daughter, Devika, whose patience, love and understanding were the major factors contributing to the successful completion of my thesis.

Kalpana

Kalpana

DECLARATION

Certified that the work incorporated in the thesis "Cloning and analysis of glutelin: A major seed storage protein in basmati rice", submitted by Mrs. Kalpana Shrikant Joshi was carried out by the candidate under my supervision. Such material as has been obtained from other sources has been duly acknowledged in the thesis.

P K Ranjekar

(P.K. Ranjekar)

Research Guide

LIST OF ABBREVIATIONS

kbp	Kilo base pairs
%	percentage
kD	Kilodalton
bp	base pairs
cm	centimeter
°C	degrees centigrade
MW	Molecular Weight
AR	analytical reagent
GR	guaranteed reagent
rpm	revolutions per minute
hr	hour
min	minute
V/V	volume by volume
ug	microgram
ml	millilitre
mg	milligram
g	gram
nm	nanometer
A	absorbance
U	units of enzyme
cv	cultivar
mA	milliampere
dATP	deoxyribose adenosine 5'-triphosphate
dGTP	deoxyribose guanosine 5'-triphosphate

dTTP	deoxyribose thymidine 5'-triphosphate
dCTP	deoxyribose cytidine 5'-triphosphate
ddATP	dideoxyribose adenosine 5'-triphosphate
ddCTP	dideoxyribose cytidine 5'-triphosphate
ddTTP	dideoxyribose thymidine 5'-triphosphate
ddGTP	dideoxyribose guanosine 5'-triphosphate
Amp	ampicillin
psi	pounds per square inch
SDS	sodium dodecyl sulphate
BSA	Bovine Serum Albumin
PEG	polyethylene glycol
B-370	Basmati-370
O/N	Overnight
RT	Room Temperature
d/W	distilled water
PCR	Polymerase Chain Reaction

LIST OF FIGURES

LIST OF TABLES

SYNOPSIS

Introduction

Rice seed proteins consist of albumin, globulin, prolamin and glutelin. Glutelin is the major seed storage protein of rice which accounts for about 80% of the total endosperm protein. Unlike zein proteins in maize, this protein is encoded by a small multigene family (about 5 copies per haploid genome) (1). It is synthesized as a large precursor (57 kD) which is then processed into two subunits by post translational cleavage (2). A few cDNA clones encoding glutelin polypeptides have been isolated and sequenced. Takaiwa et al (3) have cloned the rice glutelin gene and have determined its structure by sequencing a genomic fragment of 3.4 kbp containing the entire RNA coding region as well as 900 bp of the 5' and about 500 bp of 3' flanking regions .

All the available work on rice glutelin gene is predominantly on Japonica cultivars and very little is known about this gene in Indica rices . Since our long term goal is to carry out genetic manipulations of seed storage proteins in an indica rice, we decided to focus our attention on characterization of seed proteins and genes encoding them in an Indica variety namely Basmati-370 which is very well known for its excellent cooking qualities and aroma. In my thesis work, I have

constructed the genomic library of rice in a suitable charon vector, screened the library using an appropriate oligonucleotide probe to identify glutelin encoding clones and characterized one such clone by determining its copy number, restriction map, and genomic organization.

Thesis plan

The thesis is organised as follows :-

Chapter I: Introduction

Here, I have compiled most of the available recent literature on seed storage protein genes under the title: "Seed storage protein genes in cereals with special reference to rice".

Chapter II : Materials and Methods

During my Ph.D work, I have handled an array of molecular techniques starting from DNA extractions upto DNA sequencing. The detailed protocols of all these experimental procedures are included in this section.

Chapter III : Results

This section is divided into the following three parts:

Section I : Construction of rice (Basmati-370) genomic library in Charon 40 vector.

Section II: Screening of the rice genomic library and identification of clones containing glutelin genes.

Section III: Characterisation of rice (Basmati-370) glutelin gene.

Chapter IV : Discussion

Here, I have made an attempt to discuss the results which I have described in the three sections of chapter III.

Summary of work:-

I Construction of genomic library of rice

Since the expression of glutelin gene is strictly under specific tissue and developmental stage control, it is essential to isolate and characterize a genomic gene to elucidate its specific gene regulation. The construction of rice genomic library is the first step towards this direction. This work involves 3 major steps, namely:

(a) Isolation of a high molecular weight rice DNA and its partial digestion.

(b) Preparation of charon 40 DNA and its arms.

(c) Development of an efficient *in vitro* packaging system.

(a) Among different methods which were tried for isolation of high molecular weight rice DNA, I succeeded in isolating rice DNA of about 50 kbp size by the urea sarkosyl method (4) with a yield of 2-3 mg/10 gm of rice seedlings. Partial digestion of rice DNA was next carried out using *EcoRI* and the partially digested DNA was size fractionated on a 10-40% sucrose density gradient. The *EcoRI* digested rice DNA of size range 10-20 kbp was used for ligation with the Charon 40 vector.

(b) Bacteriophage lambda substitution vector Charon 40 (5) which can accept insert sizes of 10-20 kbp in length was used for the library construction. Charon 40 DNA was digested with *EcoRI* and *NaeI* and the two arms were purified by PEG precipitation. The Charon 40 arms and the size fractionated rice DNA were ligated using T4 DNA ligase.

(c) The development of an efficient *in vitro* packaging extract (PE) made it possible to construct the genomic library of rice (6,7). I prepared the packaging extract using a single lysogenic strain

namely E. coli SMR 10. The efficiency of packaging was 1 x 10^8 PFU/ug of DNA which was very close to the efficiency reported for this system (1.0 - 3 x 10^8 PFU/ug of wild lambda DNA).

Ligated rice DNA and Charon 40 arms were packaged using this packaging extract. The titer of the library was 2.5 x 10^6 PFU/ml. Taking into consideration the haploid DNA content of rice (1.0 pg), it appears that the library represented 8 rice genomes.

II Identification of glutelin encoding clone from the genomic library

The library was amplified before screening in order to get an over representation of the clones.

Normally cDNA clones or heterologous genomic clones are used to screen the genomic library. Since neither of these clones were available for screening the rice library for glutelin clones, I decided to construct an oligonucleotide probe from the available cDNA sequence of glutelin gene of japonica rice (8) and to use it as a probe to identify the glutelin gene. A 33 bp unique region was selected from the 5' region of the glutelin cDNA which had a very high percentage of G+C (55%) as compared to the rest of the sequence. The probe was used for plaque hybridization by end

labelling it with P-32 dATP using T4 polynucleotide kinase.

Recombinant phages were plated and plaques were lifted from the plates on nitrocellulose membranes for hybridization with the labelled oligonucleotide probe. Around 10^5 plaques were screened to obtain 13 plaques showing positive signals on autoradiography. In the second round, out of 13 plaques, only 7 gave positive signals. In further rounds of plaque lifting and hybridization, these plaques gave consistent signals which were purified at the end of 5 rounds to get seven probable clones of glutelin.

III Analysis of glutelin clones

The basic aspects of gene characterization include restriction mapping and nucleotide sequencing . I isolated DNA from the seven clones of glutelin. Out of 7 clones, 3 clones, namely lambda RG4, lambda RG6 and lambda RG7 gave strong signals on hybridization with glutelin cDNA from japonica rice. These 3 clones showed the presence of inserts of 12-15 kbp upon digestion with EcoRI. Among these clones lambda RG4 gave two insert bands of sizes 9.4 kbp and 3.55 kbp. To determine which of the 2 bands contained the glutelin gene, hybridization with glutelin cDNA was carried out. On hybridization, 3.55 kbp insert band gave signals

indicating the presence of glutelin gene in it. For further characterization of the 3.55 kbp fragment containing the glutelin gene, it was subcloned in a plasmid vector Bluescript KS+ and was named as pRG4. Restriction mapping of the clone pRG4 was carried out using different hexacutting and tetracutting restriction enzymes. When rice genomic DNA was digested with different restriction enzymes, and probed with ^{32}-P labelled pRG4 insert, a nontandem organization of the cloned glutelin gene was observed. Genomic reconstruction analysis of pRG4 insert revealed the presence of 5-8 copies of glutelin gene in basmati rice genome.

REFERENCES

1. Yamagata, H, Sugimoto, T. Tanaka, K and Kasai, Z. (1982)
   Plant Physiol., 70, 1094.
2. Zhao, M.W., Gadehouse, J.A. and Boulter, D. (1983)
   FEBS Lett., 162 96.
3. Takaiwa, F., Ebinuma, H., Kikuchi, S. and Oono, K. (1987)
   FEBS Lett., 221 43.
4. Shure, M, Wessler, S., and Federoff, N. (1983)
   Cell, 35 225.
5. Dunn, I.S. and Blattner, F.R.(1987)
   Nucl. Acids Res. 15 2677.
6. Hohn, B. and Murray,K. (1977)
   Proc. Natl. Acad. Sci. 74 3259.
7. Rosenberg, S.M., Stahl,M.M., Kobayashi,I and Stahl,F.W. (1985)
   Gene, 38 165.
8. Wang, G-S., Shastri, K., Wen, L., Huang, J-K., Sonthayanon, B., Muthukrishnan, S. and Reeck, G.R. (1987)
   FEBS Lett. 222 135.

CHAPTER I

SEED STORAGE PROTEIN GENES IN CEREALS WITH SPECIAL REFERENCE TO RICE

1.1 Introduction

The study of seed storage proteins dates as far back as 1870s with the major contribution by Osborne. Osborne's solubility fractionation method is the method of choice even today for the primary fractionation of seed proteins (1). Thus the four important classes of seed storage proteins are albumins, globulins, prolamins and glutelins which are soluble in water, salt, alcohol; and alkali, acid or detergents respectively. Of these four classes, cereals have prolamins as the most abundant proteins with rice and oats as exceptional cases where glutelins and globulins respectively are predominant. The relative seed protein content in cereals and the distribution of prolamins and glutelins are summarised in Table 1.1

TABLE 1.1

Protein content and percentage of prolamin and glutelin in cereal grains (2)

Plant	Protein %	Prolamin %	Glutelin %
Wheat	10-15	Gliadin 40-50	30-40
Maize	7-13	Zein 50-55	30-45
Sorghum	9-13	Kafirin 60	considerable
Rye	9-14	Secalin 30-50	30-50
Barley	10-16	Hordein 35-45	35-45
Rice	8-10	Oryzin 1-5	85-90
Oats	8-14	Avenin 10-15	5

1.2 Synthesis, transport and deposition of seed storage proteins

Seed storage proteins are synthesized only during the development of seeds after fertilization and are localized in the triploid starchy endosperm in cereals. The site of synthesis of seed storage proteins and site of deposition are different.

In barley, different hordeins are synthesized at different rates (3). Landry and Moureaux (1976) (4) have shown in maize that the zeins are not synthesized upto 2 weeks after pollination while glutelins are already accumulating. Gliadins of wheat are synthesized between 10-25 days after flowering (5). Oat globulin synthesis starts 3-4 weeks after flowering (6). In rice, synthesis of prolamins and glutelins is not coordinate. Glutelins are initially synthesized at 4-6 days post anthesis, whereas prolamin accumulation is first detected several days later (7).

Most major seed storage proteins are synthesized on endoplasmic reticulum (11) and then transported via golgi apparatus (12) to their site of deposition in the protein bodies. In cereals such as wheat, maize and rice, the protein bodies arise from dilations of endoplasmic reticulum (8-10).All seed storage proteins are initially synthesized with a typical leader

sequence that facilitates the transport of nascent polypeptides first through the membrane of the endoplasmic reticulum via golgi apparatus into the lumen. In golgi bodies, post translational events occur. One of the events is post translational cleavage between aspargine and glycine residues as shown in Glycine max, Phaseolus vulgaris, Pisum sativum, and Brassica napus (13). When the sequences located near the post translational cleavage site are aligned, homology is observed around putative processing site (NG). The post translational processing of rice glutelin occurs near the variable C-terminal region which is in agreement with that observed for legumes (13-15).

Protein bodies in maize, sorghum and pearl millet are ovoid organelles, embedded between starch granules (18). In wheat, barley and rye, prolamins are deposited in protein bodies which vanish at maturity in the starchy endosperm of seeds (16). At the end of seed development, clusters of proteins are deposited between starch granules that are no more surrounded by a membrane. In wheat, protein bodies associate with one another to give rise to huge clusters of proteins. These protein bodies in starchy endosperm are filled with both prolamin and glutelin (18). In rice, proteins are deposited exclusively into two morphologically

distinct protein bodies. Protein body I (PBI) stores only prolamins, while PBII is filled with glutelins. This kind of polymorphism has been observed only in rice so far among cereals and legumes (17,19).

1.3 Cloning of seed storage protein genes

To study the structure and organization of seed storage protein genes, it is necessary to clone these genes from the genome. For this purpose, genomic and cDNA libraries of plants like maize, barley, wheat, oat, rice, rye and sorghum have been constructed. In case of cDNA libraries, stage specific mRNA is isolated from which complementary DNAs are synthesized by using reverse transcriptase. This complementary DNA is cloned either into insertion vectors like lambda gt10 or expression vectors like, lambda gt11, pUC or pBS where expression is feasible. These cDNA libraries are screened using antibody probes, oligonucleotide probes or heterologous probes to identify specific cDNA clones which are mainly used to generate the sequence of the coding regions and to study expression of these genes. Table 1.2 summarises the seed storage protein genes isolated from different cDNA libraries. From the table it can be seen that most of the cDNA clones are identified by immunoscreening and hybrid released

TABLE 1.2

cDNAs encoding seed storage proteins from cereals

Plant	Seed storage protein	Vector used	Method of screening	Ref.
Wheat	gamma gliadin	pBR322	14S RNA-Probe for hybridization which translates 30,000-36,000 MW polypeptide	20
	gamma gliadin	pBR322	alpha,/β, gliadin cDNAs	21
	gamma gliadin	pBR322	Immunoscreening & hybrid selected translation	22,25
	gliadin unidentified	pBR322	Hybridization to endosperm mRNA & hybrid released translation	23
	HMW Prolamin	pUC8	Hybrid release translation	24
Rye	gamma secalin	pUC8	Hybrid selected translation	26
Sorghum	Kafirin	pUC8 gt10	-	27
Barley	B hordeins	pBR322 pPU207	Probe used-polyA+ RNA, Hybrid selected translation	28
	C hordeins	pBR322	Hybrid released translation	29
Maize	Z-19	pBR322	Probe used-Barley cDNA Hybrid releas translation	30

Plant	Seed storage protein	Vector used	Method of screening	Ref.
Maize	Z-22	pBR322	Probe used-zein cDNA Hybrid released translation	31
	Z-19/Z-22	pUC9	Zein genomic clone	32
	zein	pUC8	Immunoscreening	33
Oat	12S globulin	pUC8	Oligonucleotide probe	34
	Prolamin	λgt10	Hybrid selected mRNA translation	35
Rice	13 kD prolamin	λgt11	mRNA probes	36
	10 kD prolamin	pUC9/	rabbit antibody raised against 10 kD rice prolamin	38
		λgt11	pHIB (Partial cDNA clone from pUC9 library	38
	Glutelin	λgt10	Hybrid released translation	39
	Glutelin	pBR322	Hybrid released translation Immunopreci-pitation.	40
Rice	Glutelin	pBR325	PREE61 glutelin cDNA	41

Plant	Seed storage protein	Vector used	Method of screening	Ref.
Rice	Glutelin	λgt11	Mono specific antibodies raised against purified acidic subunit of glutelin protein	42
	Glutelin		Oligoprobe	43
	Glutelin	pBR325	Colony Hybridization & Hybrid released translation	44
	Glutelin	pBR325	Glutelin, pREE61 cDNA	45

translation. The latter involves hybridization of the cDNA to the mRNA population, release of these mRNAs from the hybrids and their in vitro translation to check if they give the desired protein.

Genomic clones encoding seed storage proteins have been isolated from the genomic libraries which are made in different lambda derived replacement vectors or cosmid vectors. Wheat (46),maize (47), sorghum (48), barley (30) and rice (49) genomic libraries have been made and seed storage protein genes encoding gliadins, zeins, kafirins, hordeins and glutelins respectively are isolated from these libraries and characterized.

1.4 Copy number and chromosomal localization of seed storage protein genes

Seed storage proteins are encoded by multigene families which exhibit a high degree of conservation and linkage among themselves. The number of genes constituting gene families has been estimated to range from 5-100. There are multiple families of cDNA clones corresponding to alpha zein mRNAs that can be distinguished from one another at different hybridization criteria (50,51,52). The cDNA clones hybridize to multiple coding sequences in maize genome leading to an estimate of 75-100 genes (53,54);β and gamma zeins, however, show presence of less than 5 copies in the genome (54,55). Unlike alpha zeins, oat globulins, wheat gliadins and rice glutelins and prolamins show presence of less than 10 copies of the gene. Hordein and gliadins cDNA clones hybridize to multiple DNA fragments in restriction enzyme digest of genomic DNA (47,56,57) demonstrating that there are multiple genes homologous to the mRNA sequences. Several zein genomic clones have been isolated in which the coding sequences have altered initiation codons or premature stop codons (58,59,60,61). These are thought to be pseudogenes but it has been suggested that transcripts from genes with premature stop codons may

be translated (52). In case of rice glutelin, three subfamilies namely Gt1,Gt2, which are closely related and Gt3 distinctly different from Gt1 and Gt2 showing significant divergence in all the portions of gene including 5' flanking sequence have been identified. Rice genome contains 5-8 copies of each of the glutelin subfamilies (42).

Seed storage protein genes occur singly or in clusters and on same chromosome as well as on different chromosomes. Little is currently known regarding the molecular organization of loci encoding wheat and barley prolamins. Recently the D subunits of the LMW glutelins of wheat have been mapped on the short arm of chromosome 1B. All the three loci controlling synthesis of barley hordeins map to chromosome 5 (26,61,62). Seven genes corresponding to zeins of MW 19 kD and 20 kD have been mapped to the short arm of chromosome 7 and linked with these are two regulatory loci Opaque-2 and De B-30 that exert qualitative and quantitative effects on zein synthesis (63). A set of nine genes corresponding to MW 20 kD and 22 kD zeins is mapped to the short and long arms of chromosome 4 and associated with these is the regulatory locus floury-2. A third locus corresponding to MW 22 kD zeins is on the long arm of chromosome 10. A gene encoding β zein (Mr

15,000) has recently been mapped on the long arm of chromosome 6. Chromosomal localization of rice and sorghum storage protein genes as well as gamma zein genes has not yet been reported.

1.5 Structural features of seed storage protein genes

The organization of seed storage protein genes is similar to that of other plant genes (64). Briefly on the 5' end of the genes upstream to the first amino acid codon, typical eukaryotic transcription initiation signals like TATA Box and CAAT Box are located. On the 3' end of the gene downstream to the translation termination codon (s) are consensus polyadenylation and transcription stop signals, AATAA. (TABLE 1.3)

Chen et al (1986) (65) have noted that 184 bp 5' region in alpha subunit of β conglycinin contains two 28 bp imperfect repeats and 5 copies of a 6 bp repeat sequence AGCCCA. This region is also found in 5' flanking region of β subunit of β conglycinin and β phaseolin in frenchbean (66). Two other repetitive elements CATGCAT (67) & AACACAA/C (68) are found in the 5' flanking sequence of the soybean glycinin, β conglycinin and seed storage protein genes from frenchbean and cotton. Similarly in cereals like barley, wheat and maize 5' flanking regions have been sequenced. 5' sequences of B1 hordein and alpha gliadins show 80% homology within 600 bp upstream of

TABLE 1.3

Structural features of seed storage protein genes in cereals

Plant	Gene family	CAAT/TATA Box (nucleotide upstream of transcription start site)	Polyadenylation signals (nucleotide downstream to translation termination codon)	References
Maize	Zeins	30 bp	occur twice, 30bp	133
Barley	Hordein	-	3 poly A signals	133
Wheat	Gliadin	TATA Box- 104 bp CAAT Box- 141 bp	2 polyA signals, 80 and 127 bp	133
Sorghum	Kafirin	TATA Box- 30bp CAAT Box- 60bp	Single consensus polyadenylation signal at 3' end	27
Oat	12S globulin	Not available	2 potential polyadenylation sequences AAT AAA	34
Rice	Glutelin	TATA Box- 28bp	Multiple polyadenylation signals AATAAA/AATAAC clustered in Gt1 & Gt2 subfamilies separated by 8 bp in Gt3	42

the cap sites. An unusual feature of some of the zein genes is the presence of multiple promoter regions separated by about 1000 nucleotides (rich in A+T bases) at their 5' end (59,69-71). In 5' flanking regions of rice glutelin gene families, -300 bp element with the consensus sequence TG (T/A/C) AAA (G/A) is present at -694,-486, and -253 bp in the promoter region. A perfect RY repeat (CATGCATG) is present at -285 bp in Gt2 glutelin gene family. Multiple polyadenylation signals, AATAAA or AATAAC are present in all the 3 glutelin gene types (41). In cereal storage protein genes, the open reading frame (coding region) is generally not interrupted by introns. The glutenin (48,72,73), gliadin (74,46,47), zeins (75) and hordein genes (48) that have been characterized so far do not contain introns in their coding sequences. In rice glutelin family, however, three introns interrupt glutelin genes at identical positions of the coding sequence. The positions of these introns are identical to those observed for pea legumin 11S storage protein gene (76). All the intron borders obey the AG/GT rule (77). The introns of the glutelin gene are relatively AT rich and all are relatively small, ranging in size from 75 to 111 bp (42). The prolamin genes of rice namely 10kD, 13kD and 16kD prolamins show absence of introns (37,78).

The presence of short introns or their total absence seems to be a prevailing feature of cereal storage protein genes (79).

A very striking observation about the storage protein genes is the presence of blocks of nucleotide sequence repetition in the coding regions such that the translated proteins also have a repeating block structure of amino acids. This repetitive block structure is presumed to have evolved in such a way so as to follow folding, aggregation and packaging of polypeptides into a specific 3-D structure which can form protein bodies.

1.6 Evolution of cereal seed storage proteins

Cereal prolamins are a heterogeneous series of proteins specified by families of polymorphic genes. Comparison of the sequences of different prolamins suggests that they are a part of a superfamily of seed proteins with a limited sequence homology (below 50%) (80). Kries *et al.* (61) have proposed a hypothesis to explain the evolution of this superfamily. According to him, the structural genes for all these proteins have evolved from a single ancestral gene through duplication and divergence. Because S-rich as well as S-poor prolamins have similar consensus repeat motifs

and homologous 3' untranslated regions (29), it is suggested that they have a common evolutionary origin. Prolamins homologous to the hordeins of barley are found in wheat (the gliadins and glutenins),and in rye (the secalins), while the major prolamins of more distantly related maize (the zeins) and sorghum (kafirin) have evolved independently (61). Rice glutelin shares several characteristics with leguminous globulins and oat 11S globulins. Comparison of rice glutelin precursor sequence with those of pea legumin A, soybean glycinin A B and rape seed cruciferin precursors reveals 37% homology at the amino acid level. When cDNA sequence of basic subunit of oat 11S globulin is compared with the corresponding region of rice glutelin, 74% homology is found between the two proteins. The regions of identity in these two proteins are dispersed throughout; the similarity is not due to convergent evolution but to divergent evolution from a common ancestral gene (42,43,44).

1.7 Regulation of seed storage protein genes

Seed storage protein genes represent a striking example of a highly regulated gene set. The gene products of such genes are present only in cell types at specific stage of development. The abundance of seed proteins and their mRNAs in the seeds have made it

possible to study the precise developmental and tissue specific regulation of their expression (81,82). Regulation of seed storage protein genes is governed by an elaborate system of cis acting control elements and trans acting gene products affecting gene expression at the transcriptional and post transcriptional level.

Cis acting transcriptional control elements are generally found in the 5' flanking region of a gene within 1-2 kbp of the transcription start site. These elements appear to be redundant, multipartite, function from either orientation and contain short repetitive oligonucleotide sequences that act as recognition sites for nuclear DNA binding proteins.

Trans acting elements code for DNA binding proteins that recognize specific cis-acting transcriptional control elements and bring about the regulation. This sequence specific DNA binding activity could be studied by techniques such as gel retardation assays and DNA footprinting. In maize, Maier et al (83) have shown that endosperm nuclei contain a DNA binding protein that recognizes a 22 bp region 5' to a maize 19 kD zein gene. A similar sequence is found 5' to the endosperm specific barley β hordein and wheat alpha gliadin genes (133).

Zein synthesis in maize is co- ordinated by at

least six trans-acting regulatory gene loci. Mutations at floury-2 (Fl2), Mucronate (Mc) and Opaque-2 (o2) and De-B30 suppress the expression of the 22 kD and 19kD zeins (84-86). Opaque 2 and 7 appear to represent independent regulatory pathways. On the other hand, Mucronate and Opaque-2 work synergistically to suppress zein synthesis (84). Hence,zein gene superfamily expression in maize endosperm is controlled by a minimum of two regulatory pathways under the combined influence of at least 6 trans-acting regulatory gene products. One hypothesis is that these genes encode a set of zein gene specific transcription factors that recognize a cis-control element common to one or more classes of zein genes. The DNA binding and regulatory functions of most transcription factors analysed to date represent functionally separate protein domains (87-89).

1.8 Expression of seed storage protein genes in transgenic plants

In recent years, gene cloning coupled with efficient systems for gene transfer has unravelled the complexity of plant gene regulation. Foreign genes could be introduced into cells from which fertile plants containing the foreign gene in all cells could be regenerated. Such transgenic plants provide a much

better background for studying gene regulation.

The first example of a storage protein gene transfer is introduction of phaseolin gene in sunflower (90). Studies on expression of phaseolin gene in sunflower and tobacco have revealed that phaseolin mRNA and protein products could be detected at high levels in seeds but not in leaves, thereby demonstrating correct organ specific expression. In bean, phaseolin is found exclusively in the protein bodies of the embryo, while in tobacco, seed storage proteins are present both in embryo and endosperm. Interestingly the transferred phaseolin gene directs the synthesis of phaseolin only in the tobacco embryo. This suggests that either the transcriptional and translational machinary of the tobacco cells is capable of discriminating between bean and tobacco regulatory sequences or alternatively that the gene may be expressed in the tobacco but the gene products may fail to accumulate in tobacco endosperm. However, the development specific expression pattern of phaseolin gene is retained in tobacco. In tobacco, the phaseolin proteins start to accumulate 15 days after pollination as in bean, whereas the synthesis of endogenous storage proteins begins only after 9 days (90,91). Similar experiments have been performed with a soybean gene coding for the alpha subunit of beta-conglycinin, which

is studied in transgenic petunia plants (92). In this case, the soybean gene is expressed predominantly in the seeds of petunia, and the timing of the expression follows the pattern of the endogenous petunia proteins.

The DNA sequences responsible for developmental regulation of prolamin and glutelin gene expression are less characterised than those for seed storage globulins because of the slower development of methods to transform and regenerate monocots. The Ti plasmid, an effective vector for dicots (101-103) transforms monocots inefficiently (93).

In monocots, the endosperm is the tissue containing most of the storage proteins. Expression of a zein genomic clone has been studied in transformed sunflower tissue using an oncogenic vector (94,95). Here transcription starts at the authentic site, but the level of expression is relatively low (1% of maize endosperm) and no protein is made. The low level of expression may not be surprising mainly for two reasons. First not all monocot genes are well expressed in dicots. Second, zeins are expressed in a highly tissue specific fashion and sunflower tumor cells may not be a suitable tissue for the expression of these genes. A zein gene, introduced in transgenic tobacco, is not expressed in seeds or in any other

tissue. Demonstration of transcriptional activity of monocot genes in the seeds of transgenic dicots may be further complicated by the fact that the endosperm, which is the major storage tissue in monocots, is only a minor component of dicotyledonous seeds (96). Perhaps genes coding for monocot storage proteins that are expressed in both endosperm and embryo such as globulins may be better expressed in transgenic dicots (97-99). B1 hordein gene of barley when coupled to CAT is expressed in a tissue and developmental stage specific manner in transgenic tobacco plants, suggesting that B1 hordein cis acting regulatory sequences are recognized by tobacco trans-acting factors (115).

Very little is known about the regulation of rice glutelin gene expression except that mRNA levels are regulated differentially during seed development (42).In a recent study, Leisy et al (100) have introduced a chimeric gene composed of Gt3 promoter and the chloramphenicol acetyl transferase gene into tobacco via Agrobacterium tumefaciens mediated transformation. The CAT activity is detected in seed extracts as early as 8 days after flowering and attains a maximum level at the onset of overall protein accumulation. 980 bp of 5' flanking sequence of Gt3 is sufficient to give the highest level of expression in

seeds of transgenic tobacco. 5' regulatory sequences of Gt2 subfamily have been identified by gel retardation and DNaseI footprint assays which show that several proteins can bind to the 5' flanking region of this cloned gene (49).

1.9 Improvement of seed storage protein quality

In cereals, there is in general an inverse relationship between the yield of the grain crop and its protein content (105). In broad terms, breeders can opt either for high yield with low protein content or a high protein content with a yield penalty. Secondly, breeding for protein quality in cereals has the disadvantage that as protein increases in amount overall nutritional quality decreases. For instance, in wheat as the protein content increases from 7% to 15%, there is a steady drop in lysine from 4% to 3%. At IRRI, a cross breeding programme has been developed for generating high protein varieties of rice with the use of some high protein indica rice varieties (106) but at the cost of much time and effort. For many crops, the time for development of an improved variety often exceeds 10 years.

The alternative strategy to improve proteins is through mutations. Opaque 2 and Floury 2 mutants of

maize have been discovered (104,107) which have elevated lysine, methionine and tryptophan. Mutants of barley such as Notch 1 and Notch 2 (108,110) and Lys 95 Lys 449 (109) and Risø mutants (110) have been screened for high lysine content. Screening of rice for this type of mutations or creating mutations artificially has been done to get the high lysine mutants of rice but so far no mutant gene of this category has been found in rice.

The ideal way of improving protein quality would be to precisely alter the nucleotide sequence of the structural genes which code for the seed storage proteins so that a seed storage protein with an improved balance of essential amino acids is obtained. Presently, recombinant DNA and gene transfer techniques in plants are opening new approaches to the improvement of the nutritional quality of plant seed proteins by direct modification of the respective genes. For example, Saalbach et al 1988 (111) have reported the introduction of 4 methionine codons into a legumin B gene of Vicia faba by changing the reading frame near 5'end and the construction of a hybrid gene containing 10 methionine codons from soybean glycinin gene and the modified legumin gene. Recently, attempts have been made to modify zein gene by two approaches namely modifications by oligo directed mutagenesis and by insertion of a

piece of a phaseolin gene containing lysine codons (112,113).

Yet another example of use of *in vitro* mutagenesis is the supplementation of a gene encoding beta phaseolin, a *Phaseolus vulgaris* storage protein. Here the number of methionine codons in the phaseolin gene is increased from three to nine by insertion of a 45 base pair (bp) synthetic duplex. Either modified or normal phaseolin genes are integrated into the genome of tobacco plants through *Agrobacterium tumefaciens* mediated transformation. Although similar levels of phaseolin RNA are detected in seeds of plants transformed with either the normal or modified (himet) gene, the quantity of himet protein is consistently much lower than normal β-phaseolin (114).

1.10 Plant transformation in monocotyledonous plants

The important aspect of characterizing and modifying the genes to get the desired characteristics is to transfer them back into the same plant and make them functional as before or in a betterway. Several attempts are being made in this direction using the reporter genes such as GUS, LUX and CAT.

Transformation of dicotyledonous plants with *Agrobacterium* is well established and has produced

stable transgenic plants expressing a number of foreign genes. This has not been the case for monocots in general. To date only a few representatives like Asparagus officinalis (116), Gladiolus (117), Onion (118), Zea mays (115) and Rice (119) have been reported to be transformable by Agrobacterium.

Since monocots are not susceptible to infection by Agrobacterium, direct gene transfer methods have been developed where a suitable construct of foreign gene is directly introduced into plant protoplasts or cell. When protoplasts are used, the genes are delivered by a variety of techniques including electroporation, heatshock and polyethylene glycol treatment. Transformation of wheat, rye and maize has been reported where above methods have been used (120).

Rice is the 1st cereal for which regeneration of whole plants from callus has been reported (121). After many years of efforts, several laboratories have also reported successful regeneration of fertile rice plants from protoplasts (122-124). All the initial reports of rice regeneration are with japonica rice but to date regeneration of fertile plants from protoplasts of indica rice is also reported (125,126). Protoplasts isolated from rice cells grown in suspension when treated with a plasmid construct carrying the `nos' promoter, the NPT II gene and CaMV

terminator region; in presence of polyethylene glycol, give about 2% transformation frequency (127). The toxicity of PEG to some protoplasts has encouraged the use of electroporation as an alternative method for DNA delivery where about 26% frequency of transformation is obtained in case of O. sativa cv. Taipei 309 protoplasts (128). Transgenic rice plants produced by electroporation mediated plasmid uptake into protoplasts have been reported (129,130).

In order to circumvent the problem of manipulating protoplasts, a method based on accelerated tungsten particles has been used to carry foreign DNA into intact cells of O.sativa. Scanty reports of transformation and regeneration of rice using other appproaches such as pollen culture of rice (131), somatic embryogenesis from mature seed and young inflorescence of wild rice (132); and Agrobacterium mediated transfer into embryo derived cultures (119) are now available.

REFERENCES

1. Osborne, T.B.(1907)
   In: "Carnegie Inst. Washington Publ., Judd and Detweiler, Washington, USA, p. 84
2. Larkins, B.A. (1983)
   In: "A Comprehensive Biochemistry, Vol. 6, (Eds. Stumpf and Conn), Academic Press, London, p. 449.
3. Bishop, L.R. (1930)
   J. Inst. Brewing, **36** 336.
4. Landry, J. and Moureaux, T. (1976)
   Qual. Plant Foods Hums. Nutr. **25**, 343.
5. Terce-Loforgue, T. and Pernollet, J.C. (1982)
   C.R. Acad. Sci. **294**, Serie III, 529.
6. Adeli, K.and Altosaar, I. (1983)
   Plant Physiol. **73**, 949
7. Yamagata, H., Sugimoto, T. Tanaka, K. and Kasai, Z. (1982)
   Plant Physiol. **70**, 1094.
8 Larkins, B.A. and Hurkman, W.J. (1978)
   Plant Physiol. **62**, 256.
9. Campbell, W.P., Lee, J.W., O'Brien, T.P. and Smart, M.G. (1981)
   Aust. J. Plant Physiol. 8 , 5.
10. Tanaka, K., Sugimoto, T., Ogawa, M. and Kasai, Z. (1980)

Agric. Biol. Chem. **44**, 1633

11. Oparka, K.J. and Harris, H. (1982)
Planta, **154**, 184.

12 Takaiwa, F., Ebinuma, H., Kikuchi, S. and Oono, K. (1987)
FEBS. Lett. **221**, 43.

13. DureIII, L.S. (1975)
Ann. Rev. Pl. Physiol. **26**, 259.

14. Chrispeels, M.J., Bollini, R. and Harris, N. (1979)
Ber. Dtsch. Bot. Ges. **92**, 535.

15. Nielsen, N.C. (1984)
Phil. Trans. R. Soc. Ser. B, **304**, 287.

16. Lycett, G.W., Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984)
Nucl. Acids Res. **12**, 4493.

17. Momma, T., Negoro, T., Hirano, H., Matsumoto, A., Udaka, K. and Fukazawa, C. (1985)
Eur. J. Biochem. **149**, 491.

18. Pernollett, J.C. and Mosse, J. (1980)
C.R. Acad. Sci., Ser. D, Sci. Nat. **290**, 267.

19. Krishnan, H.B., Franceschi, V.R. and Okita, T.W. (1986)
Planta, **169**, 471.

20. Kasarda, D.D., Okita, R.W., Bernardin, J.E., Baecker, P.A., Nimmo, C.C., Lew, E.J-L., Dietler, M.D. and Greene, F.C. (1984)
Proc. Natl. Acad. Sci. USA, **81**, 4712.

21. Okita, T.W. (1984)
Plant Mol. Biol. 3, 325.

22. Scheets, K., Rafalski, J.A., Hedgecoth, C. and Soll, D.G. (1985)
Plant Sci. Lett. 37, 221.

23. Bartels, D. and Thompson, R.D. (1983)
Nucl. Acids Res. 11, 2961.

24. Forde, J., Forde, B.G., Fry, R.P., Kreis, M., Shewry, P.R. and Miflin, B.J.(1983)
FEBS Lett. 162, 360.

25. Reeves, C.D. and Okita, T.W. (1987)
Gene, 52, 257.

26. Kreis, M., Forde, B.G., Rahman, S., Miflin, B.J. and Shewry, P.R. (1985)
J. Mol. Biol. 183, 499.

27. DeRose, R.T., Ma, D-P., Kwon, I-S., Hasnain, S.E., Klassey, R.C., Hall, T.C. (1989)
Plant Mol. Biol. 12, 245.

28. Forde, B.G., Kreis, M., Bahramian, M.B., Matthews, J.A., Miflin, B.J., Thompson, R.D., Bartels, D. and Flavell, R.B. (1981)
Nucl. Acids Res. 9, 6689.

29. Forde, B.G., Kreis, M., Williamson, M.S., Fry, R.P., Pywell, J., Shewry, P.R., Bunce, N. and Miflin, B.J. (1985)

EMBO J. 4, 9.

30. Pederson, K., Devereux, J., Wilson, D.R. and Larkins, B.A. (1982)
Cell 29, 1015.

31. Marks, M.D. and Larkins, B.A. (1982)
J. Biol. Chem. 257, 99

32. Heidecker, G. and Messing, J. (1983)
Nucl. Acids Res. 11, 4890.

33. Esen, A. (1986)
Plant Physiol. 81, 70

34. Larkins, B.A and Walburg, G.(1986)
Plant Mol. Biol. 6, 161.

35. Fabijanski, S., Chang, S.C., Dukiandjiev, S., Bahramian, M.B., Ferrara, P. and Altosaar, I. (1988)
Biochem. Physiol. Pflanzen, 183, 143.

36. Takehiro, M., Hibino, J., Kunitomo, K., Mitsukawa, N., Tanaka, K. and Fujii, S. (1990)
Mol. Gen. Genet. 221, 1.

37. Kim, W.T. and Okita, T.W. (1988)
Plant Physiol. 88, 649.

38. Takehiro, M., Shibata, D., Hibino, T., Kato, T., Kawabe, K., Takeba, G., Tanaka, K. and Fujii, S. (1989)
Plant Mol. Biol. 12, 123.

39. Takehiro, M., Kunitomo, K., Yoshimori, S., Norihiro, M., Hibino, T., Tanaka, K. and Fujii, S. (1989).
Plant Mol. Biol. 12, 723.

40. Wang, C.-S., Shastri, K., Wen, L., Huang, J.-K., Sonthayanon, B., Muthukrishnan, S. and Reeck, G.R. (1987)
FEBS Lett. 222, 135.

41. Takaiwa, F., Kikuchi, S. and Oono, K. (1986)
Mol. Gen. Genet. 208, 15.

42. Okita, T.W., Hwang, Y.S., Hnilo, J., Kim, W.T., Aryan, A.P., Larson, R. and Krishnan, H.B. (1989)
J. Biol. Chem. 264, 12573.

43. Higuchi, W. and Fukazawa, C. (1987)
Gene, 55, 243.

44. Takaiwa,F., Kikuchi,S. and Oono,K. (1989)
FEBS Lett. 206, 33.

45. Takaiwa, F., Kikuchi, S. and Oono, K. (1989)
Nucl. Acids Res. 17, 3289.

46. Sumner-Smith, M., Rafalski, J.A., Sugiyama, T., Stoll, M. and Soll, D. (1985)
Nucl. Acids Res. 13, 3905.

47. Anderson, O.D., Litts, J.C., Gautier, M-F. and Greene, F.C. (1984)
Nucl. Acids Res. 12, 8129.

48. Forde, J., Malpica, J.M., Halford, N.G., Shewry, P.R., Anderson, O.D., Greene, F.C. and Miflin, B.J. (1985)
Nucl. Acids Res. 13, 6817.

49. Kim, S.Y. and Wu, R. (1990)

Nucl. Acids Res. **18**, 6845.

50. Burr, B., Burr, F.A., St. John, T.P., Thomas, M. and Davis, R.W. (1982)
J. Mol. Biol. **154**, 33.

51. Marks, M.D. and Larkins, B.A. (1982)
J. Biol. Chem. **257**, 9976.

52. Viotti, A., Cairo, G., Vitale, A. and Sala, E. (1985)
EMBO J. **4**, 1103.

53. Hagen, G. and Rubenstein, I. (1981)
Gene, 13, 239.

54 Wilson, D.R. and Larkins, B.A. (1984)
J. Mol. Evol. **29**, 330.

55. Boronat, A., Martinez, M.C., Reina, M., Puiydomenech, P. and Palau, J. (1986)
Plant Sci. **47**, 95.

56. Kreis, M., Shewry, P.R., Forde, B.G., Rahman, S. and Miflin, B.J. (1983)
Cell, **34**, 161.

57. Shewry, P.R., Bunce, N.A.C., Kreis, M. and Forde, B.G. (1985)
Biochem. Genet. **23**, 391.

58. Spena, A., Viotti, A. and Pirrotta, V. (1983)
J. Mol. Biol. **169**,799.

59. Kridl, J.C., Vieira, J., Rubenstein, I. and Messing, J. (1984)
Gene, 28, 112.

60. Marks, M.D., Pedersen, K., Wilson, D.R., DiFonzo, N. and Larkins, B.A. (1984)
Curr. Top. Plant Biochem. Physiol. 3, 9.

61. Kreis, M., Shewry, P.R., Forde, B.G., Forde, J. and Miflin, B.J. (1985)
Oxford Surv. Plant Mol. Cell Biol. 2, 253.

62. Jackson, E.A., Holt, L.M. and Payne, P.I. (1985)
Genet. Res. **46**, 11.

63. Soave, C. and Salamini, F. (1984)
Philos. Trans. R.Soc. London Ser. B, **304**, 341.

64. Sorenson, J.C. (1984)
Advances in Genet. 21, 109.

65. Chen, Z.-L., Schuler, M.A. and Beachy, R.N. (1986)
Proc. Natl. Acad. Sci. USA, **83**, 8560.

66. Sengupta-Gopalan, C., Reichert, N.A., Barker, R.F., Hall, T.C. and Kemp, J.D. (1985)
Proc. Natl. Acad. Sci. USA, **82**, 3320.

67. Nielsen, N.C. (1985)
J. Am. Oil Chem. Soc. 62, 1680.

68. Goldberg, R.B. (1986)
Philos.Trans.R.Soc. London B. Biol. Sci. **314**, 343.

69. Langridge, P., Pintor-Toro, J.A. and Feix, G. (1982)
In: "Maize for Biological Research" (Ed. W.F.Sheridan), Plant Mol. Biol. Assoc.

70. Langridge, P. and Feix, G. (1983)

Cell,34, 1015.

71. Langridge, P., Brown, J.W.S., Pintor-Toro, J.A., Feix, G., Neuhaus, G., Neuhaus-Url, G. and Schweiger, H.G. (1985)
Eur. J. Cell Biol. 39, 257.

72. Thompson, R.D., Bartels, D. and Harberd, N.P. (1985)
Nucl. Acids Res. 13, 6833.

73. Sugiyama, T., Rafalski, A., Peterson, D. and Soll, D. (1985)
Nucl. Acids Res. 13, 8729.

74. Rafalski, J.A., Scheets, K., Metzler, M., Peterson, D.M., Hedgecoth, C. and Soll, D.G. (1984)
EMBO J. 3, 1409.

75. Rubenstein, I. (1982)
In: "Maize for Biological Research" (Ed. W.F. Sheridon),
Plant Mol. Biol. Assoc., Charlottesville, 189.

76. Lycett, G.W., Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984)
Nucl. Acids Res. 13, 6733.

77. Brown, J.W.S. (1986)
Nucl. Acids Res. 7, 1513.

78. Barbier, P. and Ishihama, A. (1990)
Plant Mol. Biol. 15, 191.

79. Messing, J., Geraghty, D., Heidecker, G., Hu, N.-T., Kridl, J. and Rubenstein, I. (1983)

In: "Genetic Engineering of Plants", (Eds. T. Kosuge, C. Meridith and A. Hollaender), Plenum Publishing Corpn. N.Y. 221.

80. Dayhoff, M.O. (1976)
Atlas of Protein Sequences and Structure, 5, Suppl. 3.

81. Dure, L.III, Greenway, S.C. and Galau, G.A. (1981)
Biochem. 20, 4162.

82. Dure, L.III (1985)
Oxf. Sur. Plant Mol. Cell Biol. 2, 179.

83. Maier, U.G., Brown, J.W.S., Toloczyki, C. and Feix, G. (1987)
EMBO J. 6, 17.

84. Salamini, F., Difonzo, N., Fornasari, E., Gentinetta, E., Reggiani, R. and Soave, C. (1983)
Theor. Appl. Genet. 65, 123.

85. Soave, C., Righetti, P.G., Lorenzoni, C., Gentinetta, E. and Salamini, F. (1976)
Maydica. 21, 61.

86. Salamini, F., Difonzo, N., Gentinetta, E. and Soave, C. (1979)
In: "Seed Protein Improvement in Cereals and Grain Legumes",
FAO-IAEA, Neuherberg, 97.

87. Brent, R. and Ptashne, M. (1985)
Cell, 43, 729.

88. Hope, I.A. and Struhl, K. (1985)
Cell, **43**, 177.

89. Hope, I.A. and Struhl, K. (1985)
Cell, **46**, 885.

90. Murai, N., Sutton, D.W., Murray, M.G., Slightom, J.L., Merlo, D.J., Reichert, N.A., Sengupta-Gopalan, C., Stock, C.A., Barker, R.F., Kemp, J.D. and Hall, T.C. (1983)
Science, 222, 476.

91. Sengupta-Gopalan, C., Reichert, N.A., Barker, R.F., Hall, T.C. and Kemp, J.D. (1985)
Proc. Natl. Acad. Sci. USA, **82**, 3320.

92. Beachy, R.N., Horsch, R.B., Rogers, S.G., Hoffmann, N.J. and Fraley, R.T. (1985)
EMBO J. **4**, 3047.

93. Hooykaas-Van Slogteren, G.M.S., Hooykaas, P.J.J. and Schilperoort, R.A. (1984)
Nature, (London), **311**, 763.

94. Goldsbrough, R.B., Gelvin,S.B. and Larkins, B.A. (1986)
Mol. Gen. Genet. **202**, 374.

95. Matzke, M.A., Susani, A., Binns, A.N., Lewis, B.D., Rubenstein, I. and Matzke, A.J.M. (1984)
EMBO J. 3, 1525.

96. Greenwood, J.S. and Chrispeels, M.J. (1985)
Plant Physiol. **79**, 65.

97. Higgins, T.V.J. (1985)
Ann. Rev. Plant Physiol. 35, 191.

98. Robert, L.S., Adeli, K. and Altosaar, I. (1985)
Plant Physiol. **78**, 812.

99. Marris, C., Gallois, P., Copley, J. and Kreis, M. (1988)
Plant Mol. Biol. **10**, 359.

100. Leisy, D.J., Hnilo, J., Zhao, Y. and Okita, T.W. (1990)
Plant Mol. Biol. **14**, 41.

101. Ream, L.W., Gordon, M.P. and Nester, E.W. (1983)
Proc. Natl. Acad. Sci. USA, **80**, 1660.

102. Horsch, R.B., Fraley, R.T., Rogers, S.G., Sanders, P.R., Lloyd, A. and Hoffmann, N. (1984)
Science, **223**, 496.

103. Deblock, M., Herrera-Estrella, L., Montagu, M.V., Schell, J. and Zambryski, P. (1984)
EMBO J. **31**, 681.

104. Mertz, E.T., Bates, L.S. and Nelson, O.E. (1964)
Science, **145**, 279

105. Grant, M.N. and McCalla, A.G. (1949)
Can. J. Res. **27**, 230.

106. Tanaka, S. (1983)
In: "Seed Proteins" (Eds. W. Gottschalk and H.P. Muller), Nijhoff. M. and Junk, W. Publishers.
EMBO J. **31**, 681.

107. Nelson, O.E., Mertz, E.T. and Bates, L.S. (1965)
Science, 150, 1469.

108. Balaravi, S.P., Bansal, H.C., Eggum, B.O. and Bhaskaran, S. (1976)
J. Sci. Fd. Agric. 27, 545.

109. DiFonzo, N. and Stanca, A.M. (1977)
Genetica Agraria. 31, 401.

110. Køie, B. and Doll, H. (1979)
In: "Seed Protein Improvement in Cereals and Grain Legumes".
IAEA Vienna, 205.

111. Saalbach, G., Jung, R., Saalbach, I. and Muntz, K. (1988)
Biochem. Physiol. Pflanzen. 183, 211.

112. Cuellar, R.E., Galili, G., Kawata, E.E. and Larkins, B.A. (1985)
Abstr.1st Internat.Congr. Plant Mol.Biol. Savannah,56.

113. Hoffman, L., Drong, R. and Donaldson, D. (1985)
Abstr. 1st Internat. Congr. Plant Mol. Biol. Savannah, 56.

114. Hoffman, L.M., Donaldson, D.D. and Herman, E.M. (1988)
Plant Mol. Biol. 11, 717.

115. Graves, A.C.F. and Goldman, S.L. (1986)
Plant Mol. Biol. 7,43

116. Bytebier, B., Deboeck, F., Degreve, H., Van-Montagu, M. and Hernalsteens, J. (1987)

Proc. Natl. Acad. Sci. USA, **84**, 5345.

117. Graves, A.C.F. and Goldman, S.L. (1987)
J. Bacteriol. **169**, 1745.

118. Dommisse, E.M., Leung, D.W.M., Shaw, M.L. and Conner,A.J (1990)
Plant Science, **69**, 249.

119. Raineri, D.M., Bottini, P., Gordon, M.P. and Nester, E.W.(1990)
Bio/Technology, **8**, 33.

120. Lorz, H., Baker, B. and Schell, J. (1985)
Mol. Gen. Genet. **199**, 178.

121. Nishi, T., Yamada, Y. and Takahashi, E. (1968)
Nature, **219**, 508.

122. Abdullah, R., Cocking, E.C. and Thompson, J.A. (1986)
Bio/Technology, **4**, 1087.

123. Toriyama, K., Hinata, K. and Sasaki, T. (1986)
Theor. Appl. Genet. **73**, 16.

124. Coulibaly, M.Y. and Demarly, Y. (1986)
Pflanzenzucht, **96**, 79.

125. Lee, L., Schroll, R.E., Grimes, H.D. and Hodges, T.K. (1989)
Planta, **178**, 325.

126. Kyozuka, J., Otoo, E. and Shimamoto, K. (1988)
Theor. Appl. Genet. **76**, 887.

127. Uchimiya, H., Fushimi, T., Hashimoto, H., Harada, H.,

Syono, K. and Sugawara, Y. (1986)
Mol. Gen. Genet. **204**, 204.

128. Zhang, H.M., Yang, H., Rech, E.L., Golds, T.J., Davis, A.S., Mulligan, B.J., Cocking, E.C. and Davey, M.R. (1988)
Plant Cell Rep. **7**, 379.

129. Shimamoto, K., Terada, R., Izawa, T. and Fujimoto, H. (1989)
Nature, 338, 274.

130. Toriyama, K., Arimoto, Y., Uchimiya, H. and Hinata, K. (1988)
Bio/Technology 6, 1072.

131. Cho, M.S. and Zapata, F.J. (1988)
Plant Sci. **58**, 239.

132. Wang, M. and Zapata, F.J. (1987).
J. Plant Physiol. **130**, 79.

133. Forde, B.G., Heyworth, A., Pywell, J. and Kreis, M. (1985)
Nucl. Acids Res. **13**, 7327.

CHAPTER II

MATERIALS AND METHODS

MATERIALS AND METHODS

2.1 Chemicals, materials and enzymes

All the chemicals used throughout the work were of Analytical Reagent (AR) or Guaranteed Reagent (GR) grade, and were obtained from Qualigens, British Drug House (BDH), Sarabhai Chemicals or E. Merck, India.

Chloroform, ethanol (Swastik Laboratory PVT Ltd, Pune) and phenol (SD's Laboratory-Chemicals Industry,Bombay) were freshly distilled prior to use. Fine chemicals such as Tris [Tris (hydroxymethyl) amino methanol], sarkosyl, SDS (sodium dodecyl sulphate) agarose, ampicillin, PEG 8000 (Polyethylene glycol MW.8000, mol. biol. grade) spermidine, and putrescine were obtained from Sigma chemical co. USA.

Molecular weight markers such as lambda DNA digested with HindIII, one kbp ladder and high molecular weight DNA marker were from Bethesda Research Laboratories (BRL) USA.

Media components namely tryptone, yeast extract and Bactoagar were from Difco Laboratories, USA. NZY broth was from Gibco, USA.

All the restriction endonucleases used for digestions of DNA were from Amersham, UK, Boehringer Mannheim (Germany), New England Biolabs (NEB), USA, or Bethesda Research Laboratories (BRL) USA.

Enzymes like DNase I, RNase A (free of DNase activity) and Proteinase K were from Sigma chemical co. USA. Polynucleotide Kinase T4 DNA ligase and calf intestinal phosphatase were from Boehringer Mannheim (Germany). Transfer membranes used for Southern blotting of DNA were either Hybond-C extra or Hybond-N from Amersham, (UK) whereas membranes for plaques and colony transfers were of NEN, Dupont USA make

Bovine serum albumin (BSA) (free of DNase activity), X-gal (5 bromo, 4 chloro 3 indolyl β D-galactoside), and IPTG (isopropyl thiogalactoside) were procurred from Boehringer Mannheim (Germany).

Radiolabelled alpha P-32-dCTP or alpha P-32-dATP or gamma P-32-dATP were from BARC, Bombay. The multiprime labelling kit from Amersham was used.

Black and white ORWO photographic film (125 ASA) and X-ray films (Kodak, Fuji) were used for gel photography and autoradiography respectively. Commercially prepared developer (IPC 163) and fixer were used for developing the X-ray films.

2.2 Seed material, growth conditions and harvest of tissue

Seeds of rice Oryza sativa (variety Basmati 370) used for isolation of DNA were obtained from National

Seeds Corporation, Pune. Seeds of japonica rice of variety, Taichung 65 were procurred from Central Rice Research Institute, Cuttack, India.

To get rice seedlings, seeds of rice (Basmati 370) were thoroughly washed with distilled water followed by washing with 50% alcohol and finally soaked overnight in distilled water. They were grown on wet cotton in dark for about 8-10 days. Rice seedlings (10-15 cm height) were cut two cm above the cotton surface and were frozen in liquid nitrogen and stored at -70°C till further use.

For harvesting leaf material, surface sterilised rice seeds were grown in pots (containing half compost and half soil mixture) (Fig 2.1). After watering for 6 to 8 weeks, leaf material was harvested and was immediately frozen in liquid nitrogen and stored at -70°C.

DNA isolated from frozen seedlings of rice (Basmati 370) was used for the construction rice genomic library. DNAs isolated from leaf tissue of rice were used in all other experiments like genomic reconstruction analysis and so on .

2.3 *E.coli* hosts and vector system used for cloning

Rice genomic library was constructed in lambda derived replacement vector Charon40. Subcloning of

FIGURE 2.1 :

Photograph of rice Basmati 370 grown in pot

FIG. 2·1

glutelin gene was done using plasmid vector Bluescript KS^+ (pBS).

The details regarding genotypes of hosts used, markers, size of vectors and restriction enzyme sites in the vectors are described in Table .1 and .2

Maintenance and storage of E. coli cultures and phage lysates

E. coli cultures were stored as stab cultures at 0-4°C for a short term. For the long term storage, 15% glycerol suspensions of the liquid cultures were kept frozen at -70°C. The stabs had to be subcultured every month.

All the lysates of bacteriophage vectors, amplified library and recombinant clones were stored as 1% chloroform stocks at 4°C for a few months without much decrease in the titer. For the storage of masterstocks, dimethylsulfoxide (DMSO) was added to the bacteriophage lysates to the final concentration of 7% v/v, mixed gently, plunged into liquid nitrogen and stored at -70°C for long term storage.

TABLE 2.1 HOST SYSTEMS USED FOR CLONING

	Host name	Genotype	Comments
1.	*E. coli* LE392	supE 44 supF 58 hsd R514 galK2 gal T22 met B1 trp R55 lac Y1.	A suppressing strain commonly used to propagat -e lambda vectors.Used as a host for Charon40 vector
2.	*E. coli* ED 8767	Sup. E44 Sup F58 hsd S3 ($^{r}B^{-}M^{-}B$) recA56 gal K2 gal T22 met B1	A recombination deficient suppressing strain used for propagation of Charon 40
3.	*E. coli* DH5 alpha	Sup E44 lac U169 (80 lac Z M15) hsd R17recA1 endA1 gyr A96 thi-1 relA1	recA-Strain used for pro -pagation of pUC18 & pBS. The lac Z M15 permits alpha complementation with amino terminus of β galactosidase encode -d in pUC vectors.
4	*E. coli* SMR10	*E. coli* C (cos 2 B xisl red 3 gam am 210 CIts 857 nin5 Sam7/)	A bacteriophage lysogen used to prepare packaging extract.(A single culture producing head and tail proteins for packaging.

TABLE 2.2 VECTOR SYSTEMS USED FOR CLONING

Vector name	size kbp	Cloning sites	Features & application
Charon40	47.6	EcoRI, SstI, XbaI, KpnI SmaI, SalI, HindIII XmaIII, NotI, AvrIII, SpeI, XhoI, ApaI, BamHI, SfiI, NaeI	1) Cloning capacity is 9.207-24.207 kbp 2) 80 copies of polystuffer which is a 235 bp repeat from Adenoviral genome. Hence easy to separate arms from stuffer region. 3) Can grow on recA$^-$ strains. 4) Selection on X-gal.
Bluescript (pBS KS+)	2.96	polylinker with 21 unique restriction sites. KpnI, DraII, ApaI, Xhol, SalI, ClaI, HindIII, EcoRV, EcoRI, PstI, SmaI, BamHI, SpeI, XbaI, NotI, EagI, BstxI, SacII, SacI	1) Polylinker flanked by T3 T7 RNA polymerase promoters for efficient *in vitro* transcription. 2) Useful for site directed mutagenesis. 3) Used for sequencing directly. 4) Selection on X-gal. 5) Good for expression studies.

Media used for the growth of *E. coli* hosts

1 NZY broth

N-Z amine	:	1%
Yeast extract	:	0.5%
NaCl	:	0.5%
$MgCl_2 7H_2O$	:	0.2%
pH	:	7.4

2 NZYM broth

NZY broth with 0.2% maltose

3 LBK medium

Bacto-peptone	:	1%
NaCl	:	1%
Yeast extract	:	1%
pH	:	7.4

4 Luria Broth (LB)

Tryptone	:	1%
Yeast extract	:	0.5%
NaCl	:	1%
pH	:	7.2-7.4

All the media were prepared in single distilled water and were autoclaved at 15 psi (pounds per square inch) for 20 min.

For the preparation of solid media 2% Difco agar was included before autoclaving and plates were poured after the mixture was cooled to 45°C. For the top

agarose, 0.7% agarose was included in the broth.

Heat labile components like antibiotics, IPTG, maltose were filter sterilized through autoclaved millipore HAWP membranes and were added freshly to cooled media (45°C).

Ampicillin : ampicillin trihydrate was dissolved at a concentration of 100 mg/ml in sterile autoclaved double distilled water. The effective concentration of 100 ug/ml was used in the medium.

X-gal : (5 bromo-4 chloro 3-indolyl beta D-galactoside) X-gal was dissolved in dimethyl formamide at 2% concentration.

IPTG : (Isopropyl thiogalactoside) 100 mM/lit.

2.4 Isolation of Rice DNA

Rice DNA was extracted essentially according to the procedure of Shure et al; 1983 (1). The frozen etiolated seedling tissue was ground in a mortar to a fine powder with 0.5 mm glass beads under liquid nitrogen. The powder was then suspended in urea-phenol extraction buffer (8.0 M urea, 0.35 M NaCl, 0.05 M Tris.HCl, pH 7.5, 0.02 M EDTA). 2% Sarkosyl and 5% Tris saturated phenol were added freshly to the buffer. The suspension was incubated at 65°C for 15 min. and then allowed to cool down to room temperature for extraction with equal volume of phenol : ($CHCl_3$: IAA) 3:1 to

remove proteins and pigments. The upper aqueous phase containing the nucleic acids was collected after centrifugation at 10,000 rpm in Sorvall RC 5B centrifuge (rotor SS34) for 10 min at 4°C and was dephenolized by extracting twice with chloroform isoamyl alcohol (24 : 1 v/v) mixture. Nucleic acids were precipitated from the aqueous layer with two volumes of chilled ethanol. The pellet was washed with 70% ethanol, air dried and suspended in TE (10 mM Tris-HCl, 1 mm EDTA, pH 8.0). RNA contaminants were removed by addition of RNase A (made DNase free by boiling for 10 min) to a final concentration of 100 ug/ml and incubating the mixture for 1 hr at 37°C. The RNase reaction was terminated by deproteinization as described above and DNA was reprecipitated from the aqueous layer by adding two volumes of chilled ethanol. DNA was spooled with a glass rod, washed twice with 70% alcohol, air dried and dissolved in TE. About 2 mg DNA was obtained from 10 gm seedling tissue.

Each DNA preparation was checked for protein and polysaccharide contamination by carrying out wavelength scanning in the range of 220-320 nm on a Shimadzu double beam spectrophotometer, model uv 210-A. Only those DNA preparations which exhibited an optical density ratio A280/A260 = 0.45 - 0.55 and A230/A260 = 0.45 and with an absorbance less than 0.05 at 300 nm

were used. The molecular weights of all the DNAs were found to be >50 kbp as determined by agarose gel electrophoresis.

2.5 Restriction enzyme digestion of DNA

a) Complete digestion of DNA

For all restriction enzyme digestions, about 3 to 4 ugs of DNA was incubated with 20 units of restriction enzyme in a reaction volume of 20 ul at 37°C for appropriate time. The enzyme to DNA and enzyme volume to reaction volume ratios were carefully adjusted to avoid non specific (star) activity of restriction enzymes (2). Following incubation, the enzyme digestion was stopped by addition of 10 X reaction terminating buffer (50% glycerol, 100mM EDTA, 0.25% bromophenol blue) to a final concentration of 1 X. Control experiments were performed using commercial lambda DNA digested with different restriction enzymes to check reaction conditions.

b) Partial digestion of rice DNA

To generate 10-20 kbp DNA fragments, rice DNA was partially digested with EcoRI restriction endonuclease (2). A solution containing 10 ug of genomic DNA with 10 ul EcoRI buffer (10X) was adjusted to final volume of

100 ul. It was dispensed in eppendorf vials with 20 ul in first vial and 10 ul each in the remaining vials kept on ice. 20 units of EcoRI were added to first vial and the contents were mixed thoroughly but gently. The tube was microfuged and 10 ul of the sample was transferred to next vial and the serial transfers were carried out up to 8th vial. The 9th vial was left as such without any addition. The vials were briefly microfuged and incubated at 37°C for 1 hr. The reaction was terminated by chilling to O°C and adding 10 X reaction terminating buffer to a final concentration of 1 X. The DNA was analysed by electrophoresis on 1% agarose, using wild type lambda DNA digested with HindIII as a standard molecular weight marker. The partially digested DNA was extracted with phenol/ chloroform and precipitated by addition of 2 volumes of ethanol after adjusting sodium acetate concentration to 0.3 M. After centrifugation, the pellet was air dried and dissolved in 50 ul TE.

c) Double Digestions of DNA

For the double digestion, DNAs were first digested with the enzyme requiring low salt assay buffer, followed by incubation of DNA with high salt requiring enzyme. Digestions were allowed to proceed at 37°C. The reaction was terminated as described above.

2.6 Agarose gel electrophoresis

All the genomic DNAs, undigested and digested, were generally analysed on 1% neutral agarose horizontal slab gels in TAE 1 X (40 mM Tris.HCl, acetate 20mM, EDTA 2 mM pH 8.1) at a constant current of 10 mA for 10-12 hr.

Phage DNA digests were analysed on 0.5% neutral agarose horizontal slab gels in TAE 1 X at a constant current of 10 mA for 12-16 hr.

Plasmid DNA digests were analysed on 0.7% neutral agarose horizontal slab gels in TAE 1 X at a constant current of 30 mA for 3-5 hr.

After electrophoresis, the agarose gels were stained in dark with ethidium bromide (100 ng/ml), visualised on a long wavelength (302 nm) UV transilluminator (UV products, San Gabriels, California USA) and photographed with a 35 mm SLR camera (Minolta x 700 with macrophotography and zoom lens system) using a red filter and ORWO 125 ASA, black and white film.

2.7 Blotting and Hybridization

a) Southern blotting

The transfer of DNA from agarose gels to Hybond N/Hybond C extra (Amersham) membranes or Nytran filters

was carried out essentially according to Southern (3) and Maniatis et al (2).

Pre treatment of agarose gels

Initially the gels were soaked in 0.25 N HCl, at room temperature for 10 min for depurination of DNA. The DNA was then denatured by soaking the gels in 1.5 M NaCl and 0.5 M NaOH for 30 min twice at RT with gentle shaking. Neutralization of the gel was done by soaking the gel in 1M Tris.HCl, pH 7.4 and 1.5 M NaCl, 30 min, twice at room temperature with gentle shaking. The mechanical/capillary transfer of DNA onto the membrane was carried out for 12-15 hr in 20 X SSC, after which the membranes were rinsed in 5 X SSC, air dried and baked for 2 hr at 80°C under vacuum (for nitrocellulose filters) or without vacuum (for nylon membranes). The filters were stored at room temperature between sheets of whatmann 3 mm paper till further use.

b) Plaque lifting

To screen the large number ($\sim 10^5$) of recombinant phages from the rice genomic library, it is necessary to transfer plaques from agar plates to the membrane for hybridization with oligonucleotide probe. Plaque blotting was carried out as described by Benton and Davis (1977), (4). The plates containing plaques were precooled at 4°C for 1.5 to 2 hr. NEN's (DUPONT) plaque

screens were placed on the precooled plates containing plaques for 4-5 min. 3 asymmetric holes were made on the paper through the plate. Membranes were gently lifted and floated on denaturation solution (0.5M NaOH, 1.5 M NaCl) for 3 min. This step was repeated again for 3 min. This was followed by 3 min equilibriation of the blot in neutralisation solution (1.0 M Tris.HCl pH 7.5, 1.5 N NaCl) twice and the blots were air dried. The blots were prepared in duplicates from each plate. For the second lift, again 3 asymmetric duplicate holes other than the 1st lift were made. All the holes were marked on the plate which were later used for aligning the signals on X-ray film with the plates.

c) Dot blotting

Approximately 10 ug DNA in 5 X SSC from each clone along with two positive controls (Basmati rice DNA (10ug) and glutelin cDNA (5)) and one negative control (Charon40 DNA), were boiled for 10 min, immediately chilled and applied on Hybond N membrane with capillary. The membrane was treated with denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 5 min followed by neutralisation solution (1.0 M Tris.HCl, pH 7.4, 1.5 M NaCl), air dried and baked at 80°C for 2 hr and were used for hybridization with glutelin cDNA.

d) Labelling of DNA

i) Multi prime reaction (6)

Probes were labelled with alpha P-32-dCTP (Specific activity 3000 Ci/m Mole by random priming method of Fienberg and Vogelstein (6) which is based on the hybridization of a mixture of all possible hexanucleotides to the DNA to be labelled, which acts as a primer for synthesis of a second strand of DNA in which P-32-dCTP is incorporated. With this straight forward method, it is possible to generate probes of specific activity > 10^8 CPM/ug DNA. The length of DNA fragments to be labelled does not influence the reaction. Since input DNA serves as a template and remains intact during the reaction, it is possible to label minimal amounts of DNA (10 ng) to a high specific activity. 100 ng of the probe DNA in a suitable volume of TE was denatured by boiling for 10 min and was immediately chilled on ice. To the denatured DNA, following reagents from Amersham multiprime DNA labelling kit were added in series: 5 ul primer, 5 ul labelling buffer, 4 ul each of 10 X dATP, 10 X dGTP, 10 X dTTP, and 5 ul of alpha P-32-dCTP (specific activity 3000 ci/m Mole) and 2 ul Klenow fragment (4U). Final volume was made up to 50 ul with sterile water, reaction was carried out at 37°C for one hr and

terminated by adding 10 X stop dye to final concentration of 1 X (10% SDS, 125 mM EDTA, 10 mg/ml stock BPB (1ml/ml).

ii) **End labelling** (7)

Oligonucleotide probes are small, single stranded DNA molecules which are synthesized chemically. These could not be labelled by multiprime reaction efficiently. 5' phosphate of the oligomer could be replaced by labelled phosphate of P-32-ATP with high specific activity.

Single stranded 33 bases long oligomer was used as a probe after labelling the 5' end of it with T4 polynucleotide kinase. The reaction was carried out by mixing the following components: Oligoprobe 1 ug, P-32-dATP 100 uci, 5.0 ul of 10X Kinase buffer (500 mM Tris.HCl pH 9.5, 100 mM $MgCl_2$, 50 mM DTT, 50% glycerol), T4 polynucleotide kinase 10U and volume was made up to 50 ul with sterile water. This was incubated at 37°C for 1 hr and then at 68°C for 10 min to stop the reaction.

e) Purification of probe

To prevent background or nonspecific hybridization in radioactive experiments, it is necessary to separate the probe from unincorporated labelled nucleotides.

This is achieved by purification of the probe by spun column chromatography (2).

To begin with a 1.5 ml microfuge tube was packed with 1ml Sephadex G-50. The entire mixture of probe, after the reaction, was pipetted onto this exclusion column. 50 ul of 25 mM EDTA/1% SDS was added directly on top of the column, and after waiting for 30 sec, the column was spun at 1000 rpm for 20-30 seconds in Remi table-top centrifuge. The probe, eluted in the volume of about 350-400 ul, was collected and used for the hybridization experiment.

Purification of the end labelled oligonucleotide probe

End labelled oligonucleotide probe was 33 bases long which would not have been excluded efficiently on any of the sephadex columns.It was therefore purified on Bio-gel P-6 column obtained from Bio-rad which has the exclusion capacity of >10 bases.

The prefilled column was spun at 2000 rpm for 2 min in table top centrifuge for packing of Biogel matrix in it. 50 µl of the reaction sample was loaded on the column, the tube was rinsed with another 50 ul of TE buffer and it was spun at 4000 rpm for 2 min. The probe was eluted in the volume of 400-500 ul which was used for hybridization.

Counting of the radioactivity

The radioactivity of the purified labelled probes was counted in Rack-beta liquid scintillation counter. The specific activity of the samples ranged between 1-5 x 10^8 cpm/ug DNA.

f) Prehybridization and hybridizations, washing of blot and autoradiography

Prehybridization

Before hybridization with labelled probe, prehybridization of the blots was carried out to block the sites on the nitrocellulose filter that bind single or double stranded DNA nonspecifically giving high background to the autoradiograms.

Southern blots or plaque screens were prehybridized for 3-4 hr at 62°C in heat sealed plastic bags containing the following mixture: 5 X SSC, 0.1% SDS, 5 X Denhardt's (1% Ficoll, 1% Polyvinyl pyrrolidone, 1% BSA), 100 ug/ml salmon sperm DNA and 5 mM sodium pyrophosphate. The amount of prehybridization solution added was 200ul/Cm^2 of the blot.

Hybridization

After prehybridization, the solution was removed

and replaced by hybridization solution containing the probe. The purified probe was denatured by boiling it together with hybridization solution for 10 min, and immediately added to the blots. Hybridizations were carried out overnight with gentle shaking. Depending on the probe used, temperature of hybridization was varied. For homologous probes, hybridization temperature was 62°C while for heterologous probes it was maintained at 55°C. Hybridizations using oligoprobes were carried out at 42°C for 12 hr and the temperature was allowed thereafter to reach room temperature (22°C).

Post-hybridization washing of filters and autoradiography

After hybridization, probe was removed and filters were washed for removal of the unbound/nonhybridizing probes using solutions of different stringencies depending on the probe used for hybridization.

The filters used for hybridization with homologous probe were washed with 2 X SSC , 0.1% SDS, 4 times each for 10 min at room temperature with gentle shaking. This was then followed by 2 washes at 62°C for 10 min each in 0.1X SSC , 0.1% SDS. In case of hybridization with oligoprobe filters were washed at room temperature for 30 min twice using 5 X SSC , 0.1% SDS followed by

two washes at 37°C with 4 X SSC , 0.1% SDS.

Moist filters were wrapped in saran wrap and exposed to X-ray film (Kodak) for 12 hr to 4 days at -70°C using Kirans superfast intensifying screens in X-ray cassetes.

2.8 Construction of rice genomic library

Construction of a genomic library of an eukaryote is an elaborate exercise involving important steps such as isolation of phage DNA and target DNA; preparation of an efficient packaging extract and annealing and ligation of vector arms and target DNA, followed by *in vitro* packaging and plating of packaged phage.

I Preparation of target DNA

Isolation of rice DNA and its partial digestion with *EcoRI* were according to the procedures described in section 2.4 and 2.5 b respectively. All partial digests of the rice DNAs were pooled together and size fractionated on sucrose density gradient (2) to get the fragments of size 10-20 kbp.

Two 10ml sucrose (10-40% w/v) gradients were formed in 12.5 ml ultracentrifuge tubes (SW41 Beckman). The sucrose solutions were made in a buffer containing 1 M NaCl, 20mM Tris.HCl (pH 8.0), 5 mM EDTA. The sample DNA (100 ug) was layered onto the gradients in total

volume of 200 ul . The gradients were centrifuged at 26,000 rpm for 16 hr at 20°C. Fractions of 250 ul each were collected from bottom of the tube. 15ul of every third fraction was mixed with 15 ul sterile water, ethanol precipitated , dissolved in 10 ul TE and was analysed by electrophoresis on 0.5% agarose gel using standard DNA markers. The fractions containing requisite size range (10-20 kbp) were pooled, dialysed against 4 litres of TE at 4°C for 16 hr with one buffer change after 6 hr and precipitated with 2 volumes of ethanol after adjusting sodium acetate concentration to 0.3M. The DNA was pelleted by centrifugation for 15 min in a microfuge and washed twice with 70% ethanol and dried. The pellet was resuspended in TE at a concentration of 1 ug/ul. The aliquot was loaded on 0.5% agarose gel to check whether the size distribution of the digestion products was in the desired range.

II Preparation of vector DNA

a) Titer checking of bacteriophage

E.coli strain LE 392 was used as a host for Charon40 phages. A fresh bacterial colony from an agar plate was inoculated into 50 ml NZY broth supplemented with 0.2% maltose and incubated overnight at 37°C. The culture was centrifuged at 5000 rpm for 10 min at 4°C.

The supernatant was discarded and pellet was suspended in 10 mM $MgsO_4$ at 1/10th of the original volume. These cells were used for transduction and the remaining cells could be stored at 4°C for 3 days for further use.

To determine the number of phage particles, the phage lysate was serially diluted 1:10, $1:10^2$, $1:10^3$, $1:10^4$ times in SM buffer (50 mM Tris HCl pH 7.4, 100 mM NaCl, 10mM $MgSO_4$,1ml 2% gelatin) and mixed with 100 ul of the above bacterial suspension. The mixture was incubated at 37°C for 30 min. for the adsorption of phages onto the bacterial cells and plated with molten NZY top agarose of temperature 45°C on NZY plates. Viable phages formed an area of clearing (plaque) zone indicating that the phages had grown and infected the bacteria. The number of plaques were counted and total plaque forming units (PFU) were determined from the dilution factor.

b) Minilysate preparation of bacteriophage from single plaque

The bacterial cell suspension (100 ul) as described in section 2.8 II a, was mixed with phage particles and incubated at 37°C for 30 min. 3 ml molten top agar was then added to it and poured on NZY plates. The plates were incubated overnight at 37°C leading to

confluent lysis. 3 ml of SM buffer was added to each of the plates and incubated further at 4°C overnight on slow shaker to allow diffusion of phages into the buffer. The bacteriophage suspension was recovered from the plate which was then washed with 1 ml SM buffer. To this suspension, a few drops of chloroform were added and it was then centrifuged in a microfuge. The supernatant was stored at 4°C with 1% chloroform which further served as a source for phage stock. The plate lysates were also used for minipreparation of phage DNA.

c Preparation of large scale lysate (2,8)

Cell density of the bacterial cell suspension which was made as described in 2.8 II a, was estimated turbidometrically at 600 nm (1 OD_{600} = 8 x 10^8 cells/ml). The cell suspension containing 1.5 x 10^8 cells was diluted to 3 ml using NZY broth. To this, 3 ml of 0.01 M $CaCl_2$, $MgCl_2$ solution ($MgCl_2$ is required for the phage stability and $CaCl_2$ causes some changes in the cell wall of the host rendering better phage adsorption) and 3.0 ml phage lysate (2 x 10^8 PFU/1ml) was added. This infection mixture was then incubated at 37°C for 20 min for the phage adsorption to take place. The mixture was transferred to a 500 ml Erlenmeyer flask containing 100 ml prewarmed NZY broth. Multiple

flasks (8) were prepared in a similar manner and incubated at 37°C on shaker at 180 rpm to get the bulk lysate. After 4-5 hr, complete lysis of cells was observed, 1% chloroform was added to each of these flasks and they were further kept on shaker for 15 min at 37°C to ensure complete lysis. Lysates were then centrifuged at 10K rpm in a Sorvall GSA rotor at 4°C for 20 min to remove the debris. The supernatant was collected and phage titer was determined as mentioned in part (a), of this section.

d) Extraction of DNA from bacteriophage lysate

The lysate obtained as described above was incubated in presence of DNase I (1 ug/ml of lysate) and RNaseA (5 ug/ml of lysate) at 37°C for 1 hr to remove host DNA and RNA respectively. Phage was then precipitated using 5.8% NaCl and 10% polyethylene glycol-6000 (PEG) at 0°C overnight as described by Yamamoto et al (1970) (9). The precipitated phage was pelleted at 10K rpm for 30 min at 4°C in a Sorvall GSA rotor. Supernatant was decanted and phage pellet was resuspended in SM buffer. Traces of PEG in the pellet were removed by extraction with chloroform. Concentrated phage was then layered on CsCl step gradient (3.5 ml, d = 1.7; 2.5 ml, d = 1.5; 2.5 ml, d = 1.3) in SW 27 centrifuge tube and rest of the tube was

filled with SM. This was spun in SW27 rotor for 2 hr at 24,000 rpm, 4°C. Bluish band of phage was removed with a syringe (25 G needle). Dialysis was then carried out in SM buffer at 4°C, twice changing the buffer. Purified phage was then lysed in presence of EDTA (final concentration 20 mM), SDS (final concentration 0.5%) and proteinase K (50 ug/ml) at 68°C for 15 min. Phage DNA was then extracted once with equal volume of buffer saturated phenol: Chloroform : IAA (25 : 24 : 1) and twice with chloroform : IAA (24:1). Aqueous phase was chilled and DNA was precipitated by adding twice the volume of chilled ethanol. For completion of DNA precipitation, this was kept at -20°C for 1 hr. Phage DNA was centrifuged down, supernatant decanted and pellet washed with 70% ethanol. Pellet was then dried and dissolved in TE buffer. The purity and quality of DNA were checked spectrophotometrically from the absorbance at 230, 260, 280 and 310 nm. Also the DNA was checked by electrophoresis in 0.5% agarose gel.

e) Isolation of arms of Charon40 vector by PEG precipitation

Lambda derived Charon40 vector can accept large inserts of 10 kbp to 20 kbp. It has got 16 restriction sites in its inverted multiple cloning site. EcoRI is the outermost restriction site separating the stuffer

from the arms. The stuffer region is composed of 80 copies of 235 bp long repeat of adenoviral origin which could be separated by digestion with NaeI restriction enzyme. These 235 bp stuffer fragments can then be removed selectively by precipitation with PEG from high molecular weight annealed arms (left arm :19.2 kbp, right arm :9.6 kbp, annealed arms : 28.8 kbp) (10)

i) Digestion of Charon 40 DNA with EcoRI and NaeI

100 ug of Charon40 DNA was digested with NaeI (10 u/ug DNA) at 37°C overnight. To this NaeI digested DNA, EcoRI (10 u/ug DNA) was added and digestion was continued for 5 hr. The digestion was checked on 0.5% agarose gel, electrophoresing 1 ug of digested DNA along with the control and marker DNA. After assuring complete digestion with EcoRI as well as NaeI separating the two arms and the stuffer, DNA was extracted twice with phenol/chloroform/IAA (25:24:1) and chloroform/IAA (24:1) and precipitated with 2 volumes of ethanol in presence of 0.3 M sodium acetate concentration. This was kept at -20°C for 30 min to get complete precipitation of DNA.

ii) Annealing of phage arms

After centrifugation the pellet was dried and dissolved in 875 ul of TEN buffer (0.57 M NaCl, 11.4 mM

Tris.HCl pH 7.4, 1.2 mM EDTA). To this sample $MgCl_2$ was added to final concentration of 10 mM. This mixture was heated to 68°C for 10 min and then incubated at 42°C for 1 hr followed by cooling to room temperature. This facilitated annealing of the cos ends of phage.

iii) Precipitation of phage arms by PEG

To 875 ul of annealed and digested charon40 DNA 125 ul 40% PEG (in water) was added, mixed and kept for precipitation for 2 hr at 22°C. This sample was spun for 20 min in microfuge. Supernatant was decanted and the pellet was dissolved in 500 ul of 0.2 M NaCl, and precipitated with 1 ml ethanol at -20°C O/N. After spinning in microfuge, DNA pellet was dried and dissolved in 20ul TE.1 ul of this sample was diluted to 50 ul for wavelength scanning as described in section 2.4. Purified arms were checked on 0.5% agarose gel.

III Preparation of packaging extract

Regeneration of phage particles from purified recombinant lambda DNA is an important step in preparation of genomic library in lambda derived vectors. Different methods developed for in vitro packaging of lambda make use of pairs of lysogens that have complementary defects in lambda packaging protein. When these two lysogens are combined, both added lambda

DNA and endogenous lambda DNA are packaged by the full complement of packaging proteins (11). We have prepared packaging extract using the latest method of Rosenberg et al (12,13). A single strain of E. coli SMR10 coding for all the packaging proteins but unable to package endogenous phage DNA because of the cos2 mutation in the packaging origin was used for this purpose.

Preparation of inoculum

A successful preparation of packaging extracts is dependent on the specific mutations in bacteriophage lambda genome, so the growth and maintainance of this lysogen is very critical.

Master stocks of E. coli SMR 10 were stored in -70°C as glycerol stocks. To start with 2 ml LB medium was inoculated from glycerol stock and incubated at 32°C O/N on a shaker at 180 rpm. From this liquid culture, E.coli SMR10 was streaked on LB agar plates and incubated at 32°C O/N. The presence of the mutation that renders the cI gene product temperature sensitive could be assured by checking the inducibility at 45°C. Colonies were picked up and replica plated on two LB agar plates. One plate was incubated at 32°C and another at 45°C. One of the colonies which did not grow at 45°C but grew well at 32°C was used for inoculating 20 ml liquid LBK broth at 32°C, overnight.

Preparation of packaging extract

For preparation of packaging extract, the lysogenic bacteria are first grown at 30-32°C to midlog phase, lytic functions are then induced by inactivating the cI repressor protein at 45°C, and finally the cultures are grown for an additional 2-3 hr at 38-39°C to allow packaging components to accumulate. Cell extracts are then prepared from these induced cells. 115 ml LBK broth, prewarmed to 34°C in each of four 2 lit flasks, was inoculated by approximately 6 ml of O/N grown E.coli SMR 10 culture to give an initial OD600 of 0.015. Cells were grown at 34°C with vigorous (300 rpm) agitation until the OD_{600} reached 0.8 (2.5 - 3 hr). At this OD, culture was induced by warming up to 45°C for 15 min with vigorous shaking. After induction of culture, flasks were incubated for 90 min at 37°C with vigorous agitation for accumulation of packaging proteins.

Cultures were then quickly chilled by swirling the flasks in ice water for 5 min, transferred to two precooled 250 ml centrifuge bottles and centrifuged at 4000g for 6 min at 4°C in precooled rotor to collect the pellet.

4.5 ml precooled TSP (40 mM Tris.HCl pH 7.9, 10 mM spermidine 3 HCl, 10 mM putrescine 2 HCl) was added to

each of the pellet. Spermidine stabilizes the phage structure and putrescine allows packaging of DNA molecules of a greater size range to be packaged. Bacterial pellets were suspended in TSP using chilled pasteur pipette and transferred to 12 ml centrifuge tube. After centrifugation at 4000g for 6 min at 4°C in a precooled rotor, the pellet was suspended in 0.35 ml of precooled TSP using chilled thin glass rod.

20 ul of above suspension was distributed in the microfuge tubes containing 5 ul DMSO + ATP mixture (50% DMSO, 7.5 mM ATP,pH 7.0), tubes were vortexed and frozen in liquid nitrogen and stored at -70°C.

IV Annealing and ligation of vector arms and rice DNA

Annealing of purified arms was done as described in section 2.8 part II e.

Theoretically, 2 molecules of arms ligate with 1 molecule of insert to give one recombinant molecule. From these theoretical considerations, it is possible to predict the optimal concentration of vector and insert. However, because of the difficulty of estimating small quantities of DNA and the inability of some ends to ligate, it is helpful to conduct a series of test ligations to determine the ratio of reagents that will result in the maximum number of clones.

Annealed Charon40 arms and EcoRI digested rice DNA

fragments in the size range of 10-20 kbp were mixed in the following proportion to achive the ratios of 2:1, 1:1, 0.5:1 (Charon40 arms : rice DNA). In each case, 900 ng of total DNA was mixed with 1/10th volume of 10X ligation buffer (500mM Tris.HCl, pH 8, 70 mM $MgCl_2$, 10mM DTT,10mM ATP) and incubated overnight at 14°C after addition of T4 DNA ligase (10 units) including a control test in which no insert DNA was added. The ligatability of Charon40 arms and 10-20 kbp rice DNA was checked by 0.5% agarose gel electrophoresis. Ligated DNA was then stored at 4°C till packaging.

V In vitro packaging and plating of packaged phage

In vitro packaging of all the ligation mixes in different ratios was done using freshly prepared packaging extracts. Parallel to actuall experiment, experiments involving positive and negative controls were set up. In the positive control, wild type lambda DNA, Charon40 DNA and ligated arms were packaged. In the negative control no DNA was added. To the ligated DNA, 10 X TEK (100 mM Tris.HCl pH 8.1, 5mM EDTA pH 8.0 5 mM KCl) was added to make final concentration of 1X.

Packaging extract tube was removed from -70°C and placed on ice. Immediately 10 ul DNA in 1 X TEK was added to the tube and mixed with thin glass rods. Mixing and thawing of the extract took place

simultaneously. Only one tube was processed at a time to avoid thawing of the tubes prior to mixing. This was followed by incubation at 28°C for 90 min. At the end of 90 min, 500 ul SMCK buffer (SMCK A:8 ml + SMCK B:2ml + sterile d/w 190 ml).

SMCK A: Na_2HPO_4 1.75 gm, KH_2PO_4 0.75 gm, d/w to make up vol to 10 ml.

SMCK B: NaCl 0.5 gm, NH_4Cl 1.0 gm, 1 M $MgCl_2$ 1.0 ml,1M $CaCl_2$ 0.1 ml.)

containing 50 ug/ml DNaseI was added to each tube and mixed well. DNase I digests the released E.coli DNA and unpackaged phage DNA. After 5 min,100 ul of chloroform was added to the tubes which were then vortexed and microfuged.

Plating of in vitro packaged phage

The infection mixture was prepared by 10 fold serial dilutions of packaged phage in SM buffer, plating bacteria and 10 mM CaMg solution (10% or more packaged phage in infection mixture causes inhibition of adsorption). Plating was done as described earlier in section 2.8 II a. Recombinant phages obtained on packaging the ligated DNA represent the rice genomic library.

VI Amplification and storage of the library

The library of recombinant bacteriophages is amplified by growing a plate stock directly from the packaging mixture as described in section 2.8 II b.

Amplified library in the form of a lysate was stored at 4°C with 0.3% chloroform without reduction in the titer.

2.9 Identification of clones containing glutelin gene from the amplified rice genomic library in Charon40 vector

Screening of the rice genomic library to isolate glutelin coding clones was done using an oligonucleotide probe prepared from the published glutelin cDNA sequence from a rice japonica variety (14). For this purpose, amplified rice genomic library was plated on NZY agar plates of size 150 mm to get approximately 10,000 PFU/plate. Plaque lifts of total 10 plates were made as described in section 2.6 C.

Hybridization with 33 bp long end labelled oligonucleotide probe was carried out as mentioned in section 2.6. On autoradiography, the signals were matched with the plates and the regions on the plates showing positive signals were removed and suspended in SM buffer. Phages were diluted and again plated on NZY

agar plates to get around 1000 plaques per plate and hybridizations were carried out as described above. Again the plaques showing positive signals were picked up and suspended in SM buffer. Plating on NZY agar plates was carried out to get about 100 PFU/Plate. The above steps were repeated till all the 100 plaques in the plate showed positive signals on autoradiography.

In further rounds of plaque lifting and hybridization with the oligoprobe and autoradiography seven plaques gave positive signals consistantly. Positive clones were named as lambda RG1 to lambda RG7.

2.10 Subcloning of glutelin gene in a plasmid vector

a) Preparation of vector DNA

Plasmid Bluescript KS+ (22) DNA was digested with EcoRI according to section 2.5 a. The digestion was checked on 0.7% agarose gel to ensure complete linearization. The digested plasmid DNA was deproteinized prior to it's dephosphorylation.

In order to hydrolyze 5' phosphate from the linearized pBS, to prevent self-ligation during ligation, treatment with calf intestinal phosphatase (CIP) is essential. Dephosphorylation was carried out by setting following reaction: Bluescript DNA 200ng, 10

X CIP buffer 2 ul (0.5 M Tris.HCl, pH 9.0 ; $MgCl_2$ 10 mM, $ZnCl_2$ 1 mM, spermidine 10 mM), CIP 1U, total volume was made up to 20 ul with sterile water. Above reaction was incubated at 37°C for 30 min followed by 10 min at 75°C. After deproteinization with phenol/chloroform/IAA (3:1) mixture followed by chloroform/IAA the DNA was precipitated from the aqueous phase with 2 volumes of chilled ethanol. The precipitate was spun, dried and dissolved in T_{10}.

b) Preparation of insert DNA

Lambda RG4 DNA was isolated in the same way as mentioned in section 2.8 II and digested to completion using EcoRI as described in section 2.5 a. The DNA was loaded on 0.7% agarose gel and 3.55 kbp band which gives positive signal after hybridization with glutelin cDNA was cut from the gel and added to ultrafree - MC filter unit (Durapore 0.45 um). This was frozen in -70°C for 1 hr and thawed at 37°C for 10 min. Unit was centrifuged at 10,000 rpm for 10 min. To the agarose pieces on top, 100 ul TE was again added and column was again spun at 10K for 10 min. Filtrate collected in the tube was extracted with phenol/chloroform and chloroform/IAA and aqueous layer was precipitated by adding two volumes of ethanol after adjusting sodium acetate concentration to 0.3 M. DNA was dissolved in

sterile water. Eluted insert was checked on 0.7% agarose gel.

c) Ligation of 3.55 kbp insert containing glutelin gene with plasmid Bluescript DNA

3.55 kbp insert DNA (100 ng) and EcoRI digested CIP treated Bluescript DNA (50 ng/ul) were mixed in a reaction volume of 20 ul. Reaction mixture contained 2 ul 10 X ligation buffer and 2 U T4 DNA ligase. Ligation reaction was carried out at 12°C for about 12-16 hr. This ligation product was directly used to transform E.coli DH5α competent cells.

d) Preparation of competent cells

Although different methods are available for bacterial transformations, the procedure developed by Hanahan (1983) is, however, routinely used (15). This procedure can yield competent cultures of E. coli strain DH', DH5, DH5 and NM294 that can be transformed at frequencies more than 5 x 10^8 transformed colonies per microgram of plasmid DNA.

A single colony of E. coli DH5α was inoculated in 5ml LB and was grown overnight at 37°C with shaking (250 rpm). 1 ml of this saturated overnight grown culture was inoculated in 50 ml LB; and was grown at 37°C with shaking (250 rpm), till the culture obtained

an OD_{550nm} of 0.45 - 0.55 (cell density of 5 x 10^7 cells/1 ml). Subsequently, the culture was immediately chilled on ice for 20 min to stop the cells dividing further. Cells were pelleted by centrifugation at 5000 rpm for 10 min at 4°C in a Sorvall RC5B centrifuge. The cell pellet was gently suspended in 25 ml ice cold 100 mM $CaCl_2$ and 10 mM Tris.HCl pH 7.5 solution and stored on ice for 20 min. Exposure to calcium ions renders the cells permeable to take up DNA. Cell suspension was again centrifuged for 5 min at 2,500 rpm, 4°C and the pellet was suspended in 2 ml ice cold $CaCl_2$ solution. 200 ul of competent cells were then dispensed into prechilled sterile propylene tubes. At this stage, cells are considered to be competent i.e. ready to take up the foreign DNA. Such competent cells can be stored at -70°C with 15% glycerol for use.

e) Transformation of E. coli DH5α

Transformation of E. coli DH5α was carried out with 50-100ng of ligated DNA. Along with the ligated DNA, 25 ng of uncut plasmid DNA and host DH5α cells were used as positive and negative controls, respectively.

The ligated and control plasmid DNAs were added to the cell aliquots and stored on ice for 40 min. Cells were then heat shocked at 42°C for 2.5 min. Heat shock

treatment enlarges the cell pore size allowing the DNA to enter the cells efficiently. Immediately after this step, 800 ul of sterile LB medium was added to each tube, and tubes were incubated at 37°C for 1 hr. This helps the host cells to recover and express antibiotic resistance. To this culture, 120 ul of X-gal (20 mg/ml) was added, mixed well and 100 ul aliquots of this transformation culture were spread on hard Luria agar plates (Agar 1.5%) containing ampicillin (100 ug/ml). When the plates were dry, they were incubated at 37°C, for 12-16 hr.

f) Selection of recombinants and their storage

Colonies of E. coli, harboring plasmids with inserts (chimeric plasmids), were selected on the X-gal, ampicillin LB plates. White colonies represented the recombinants, while blue were the nonrecombinants. The positive control (pBS DNA alone) showed all blue colonies; whereas the tube containing no DNA did not show a single colony on ampicillin plate. Individual white colonies were then inoculated with sterile toothpicks on masterplates containing LB agar with 100 ug/ml ampicillin.

2.11 Restriction mapping and DNA sequencing of Glutelin gene

1) Isolation of plasmid DNA

All plasmid DNA extractions were done by the alkaline lysis protocol described by Birnboim et al 1979 (16).

A single recombinant colony from the master plate was inoculated in 1 ml LB containing 100 ug/ml ampicillin. This was grown to saturation at 37°C, overnight with shaking at 250 rpm. The culture was spun in 1.5 ml eppendorf vial for 1 min, cells were pelleted, and supernatant was discarded. Cell pellet was suspended in 100 ul solution I (GTE buffer: 50 mM glucose, 25 mM Tris.HCl, pH 8.0, 10 mM EDTA) vortexed and incubated at room temperature for 5 min. Then 200 ul of freshly prepared solution II (1% SDS, 0.2 N NaOH) was added to the cell suspension, mixed well by tapping the tube, and the tube was allowed to rest on ice for 5 min. Here the cells are lysed and the bacterial proteins and chromosomal DNA get denatured. The suspension was neutralized by adding 150 ul potassium acetate buffer (solution III : 60 ml 5M acetate, 11.5 ml glacial acetic acid, final volume 100 ml with sterile water) and was again stored on ice for 10 min.

This was spun for 5 min in microfuge to pellet out the cell debris and chromosomal DNA. The reannealed plasmid supernatant was deproteinized by phenol/chloroform/IAA and chloroform/IAA treatment. The supernatant was transferred to a sterile tube and plasmid DNA was precipitated by ethanol. The DNA was pelleted by spinning and the pellet was air dried and dissolved in a suitable amount of TE.

For the plasmid midi and maxi preparations, the volumes of the buffers (solution I, II, III) were approximately scaled up.

Recombinant plasmid pRG4 DNA, isolated by above method, gives much cleaner DNA preparation that can be used further for restriction mapping analysis.

2) Restriction mapping of pRG4

The size of the insert DNA of pRG4 was determined by digesting the plasmid DNA with EcoRI and analysing the digest by agarose gel electrophoresis. Restriction mapping of pRG4 was done by subjecting the plasmid DNA to single or double digestion with specific restriction enzymes.

3) Genomic reconstruction of glutelin gene

Total rice (Basmati 370) DNA (10 ug) was digested to completion with the restriction enzymes EcoRI,

BamHI, ,HindIII and HinfI followed by electrophoresis on 1% agarose gel . The pRG4 insert was co-electrophoresed in mixtures that represented 1,2,5,10 and 20 copies per haploid genome equivalent. Following formula was used to calculate the copy number of pRG4 insert per haploid genome.

$$M = \frac{n \times D \times P \times 660}{K \times 6.022 \times 10^{23}}$$

where
n = no of gene copies = 1 copy
D = 10 ug rice DNA per lane
P = 3550 bp glutelin gene size.
K = rice haploid DNA content
= 1 pg. (19)

M = 38.3 pg.

38.3 pg of 3550 bp probe will make 1 copy in one haploid rice genome.

The DNA was transferred onto Hybond N membrane by method of Southern transfer described in section 2.7 a. pRG4 insert was used as a probe during hybridization after labelling it with multiprime reaction.

The gene copy numbers were estimated by densitometric comparison of the hybridization signals in genomic DNA to those obtained from the reconstruction standards. The gene copy number was determined from the reconstruction blots based upon a haploid genome content of rice (B-370) being 1.0 pg.

4) Sequencing of glutelin gene

The cloned rice glutelin gene, isolated from the purified recombinant plasmid pRG4, was sequenced essentially according to Sanger et al (18).

a Preparation of template DNA

For the sequencing reaction, it is necessary to have the recombinant plasmid DNA in a single stranded form. Denaturation of pRG4 was done by method of Chen and Seeburg, 1985 (21).

2 ug of plasmid DNA to be sequenced was taken in a microfuge tube and dried. To this, 20 ul of denaturation mix (0.2 M NaOH, 0.2 mM EDTA) was added and the sample was left at RT for 5 min. 2 ul 2M ammonium acetate pH 4.5 was added to neutralize the sample. DNA was precipitated by 50 ul of cold ethanol at -70°C for 30 min. The DNA was pelleted at 10,000 rpm for 10 min and the pellet was washed once with 70% ethanol, dried and dissolved in 7 ul of sterile water. This single stranded template was further used for primer annealing and sequencing.

b Dideoxy chain terminating sequencing reaction

The cloned rice glutelin gene was sequenced essentialy according to Sanger et al (18).

c Casting and electrophoresis of sequencing gel

Cleaning of glass plates

Before pouring the gel, the glass plates were thoroughly cleaned sequentially with tap water, distilled water and 70% ethanol. Then only the notched plate was siliconised using about 0.5 ml of 2% Silane (dichlorodimethyl silane) to ensure that the gel did not stick to it after electrophoresis.

Preparation of denaturing polyacrylamide gels and electrophoresis

The spacers of 0.4 mm thickness were placed on siliconised plate. The smooth plate was aligned on it properly and the bottom and the sides (about 3-4 inches) of both the plates were taped and clamped.

The 6% polyacrylamide gel mix was prepared by mixing 16 ml of 40% acrylamide solution (38 g acrylamide and 2 g bis-acrylamide in 100 ml volume), 10 ml 10 X TBE buffer (108 g Tris, 55 g Boric acid, 9.3 g EDTA/lit) and 42 g urea. The solution was heated slightly to dissolve urea and then the volume was made up to 100 ml. The solution was filtered , deaerated for 5 min and was stored at 4°C in a dark coloured bottle till further use. Before pouring the gel, 80ul TEMED (N,N,N',N',-tetramethyl-ethylenediamine) and 800 ul of

10% ammonium persulphate were added to 100 ml of 6% polyacrylamide gel mix. After pouring, the comb was immediately inserted and the gel was allowed to set for atleast 30 min. Then, the tapes were removed and the gel was clamped to the electrophoresis unit. The unit and the glass plates were maintained at a constant temperature of about 45-50°C using a circulating water bath. After pre electrophoresis at 800 V for 30 min, the samples were loaded and electrophoresed at 2000 V till the bromophenol blue (BPB) ran out. Immediately the sample was loaded again and electrophoresis was continued till BPB just touched the lower edge of the gel. This differential migration helped to read the sequence at proper base positions. TBE buffer 0.5 X was used as a running buffer during electrophoresis.

d Autoradiography and sequence analysis

After electrophoresis, the plates were dismantled in such a way that the gel remained evenly stuck to the smooth plate. The gel with the glass plate was wrapped in a saran wrap and was exposed to X-ray film for about 16 hr at -70°C. The film was then developed as described in section 2.6 f.

Restriction enzyme site analysis, search for open reading frame and comparison of the partial sequence were carried out using SEQAID II program of D. Jones

which was a kind gift by Prof. Muthukrishnan, Kansas State University, Kansas, USA.

REFERENCES

1. Shure, M., Wessler, S., Federoff, N. (1983)
   Cell, 35 225
2. Sambrook J., Fritsch, E.F. and Maniatis, T. (1989)
   In: Molecular Cloning : A Laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory Press, New York.
3. Southern, E.M. (1975)
   J. Mol. Biol., 98 503
4. Benton, W.D. and Davis, R.W. (1977)
   Science, 196 180
5. Co-Shine Wang., Shastri, K., Wen, L., Huang, J.K., Sonthayanon, B., Muthukrishnan, S., Reeck, G.R. (1987)
   FEBS Lett.,222 135
6. Feinberg, A.P., Vogelstein, B. (1983)
   Anal. Biochem. 132 6
7. Maxam A.M and Gilbert, W. (1977)
   Proc. Natl. Acad. Sci.,74 560
8. Blattner, F.R., Willams, B.G., Blechl, A.E., Denniston-Thompson, K., Feber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Sheldon, E.L., Smithies, O. (1977).
   Science,196 161
9. Yamamoto, K.R., Alberts B.M., Benzinger, R., Lawhorne, L., and Treiber, G. (1970)

Virology,40 734

10. Blattner, F.R., Dunn, I.S. (1987)
Nucleic Acids Res., 15 2677

11. Sternberg, N.D. Tiemier, and Enquist L. (1977)
Gene, 1 255

12. Rosenberg, S.M., Stahl, M.M., Koboyashi, I., Stahl, F.W. (1985)
Gene 38 165

13. Rosenberg, S.M. (1985)
Gene, 39 313

14. Takaiwa, F., Kikuchi, S., Oono, K. (1987)
Mol. Gen. Genet., 208 15

15. Hanahan, D. (1983)
J. Mol. Biol., 166 557

16. Birnboim, H.C. and Doly, J. (1979)
Nucleic Acids Res., 9 2989

17. Bennett, M.D., Smith, J.B., Heslop Harrison, J.S (1982)
Proc. R. Soc. Lond, B216 179

18. Sanger, F., Nicklen, S., and Coulson, A.R. (1977)
Proc. Natl. Acad. Sci. (USA)., 74 5463

19. Chen, E.Y., Seeburg, P.H. (1985)
DNA, 4 165

20. Short, J.M., Fernandez, J.M., Sorge J.A., Huse, W.D. (1988)
Nucleic Acids Res, 16 7583.

CHAPTER III

RESULTS

SECTION I

Construction of rice (Basmati-370) genomic library in Charon 40 vector

Introduction

Genomic and cDNA libraries of rice have been constructed predominantly from japonica cultivars. Very little work is done on indica cultivars which are widely consumed in India and neighbouring countries. I was interested in isolating and characterizing glutelin gene (the major seed storage protein of rice) from indica rice. My main objective was, therefore, to construct the genomic library of Indica variety of rice namely Basmati 370 which is very well known for its excellent cooking quality and aroma.

I have constructed a complete genomic library of rice in a lambda derived replacement vector, Charon40, which is known to be advantageous for cloning of plant genes.

3.1.1 Isolation of high molecular weight rice DNA

The size of the insert DNA to be successfully retained by the lambda Charon40 arms in a recombinant molecule has been reported to be between 10-20 kbp. To clone a foreign DNA into the lambda vector, the DNA

requires enzymatic manipulations to generate 15-20 kbp fragments which have cohesive ends compatible with those of the vector arms. Thus the original DNA prepared should be of sufficient high molecular weight (> 50 kbp). In the present work, rice DNA of size more than 50 kbp was prepared by the method of Shure et al (1) (Fig.3.1). About 3-4 mg DNA per 50 gm seedlings tissue was obtained. The digestibility of this DNA was checked using different restriction enzymes. From Fig.3.1, it can be seen that the DNA is digested to completion with a number of hexacutting restriction enzymes such as EcoRI, HindIII and BamHI.

3.1.2 Partial digestion of rice DNA with EcoRI

Since the cloning capacity of Charon40 is 10-20 kbp, it was necessary to standardize partial digestion conditions to obtain rice DNA fragments in this size range. The DNA was therefore digested with different units of EcoRI keeping time of incubation constant. After digestion, the DNA samples were checked on 0.5% agarose gel along with marker DNA, (Fig.3.2A). For preparative isolation of partially digested rice DNA, 100 ug of DNA was digested with different units of EcoRI at 37°C for thirty minutes and fractions yielding a maximum proportion of 10-20 kbp fragments were pooled together for further work.

FIGURE 3.1 : **Electrophoresis of high molecular weight rice DNA and restriction enzyme digested rice DNA**

Lane a : Undigested lambda DNA
Lane b : Lambda HindIII digest as MW marker
Lane c : Rice (Basmati 370) DNA
Lane d : EcoRI digest
Lane e : HindIII digest
Lane f : BamHI digest

Electrophoresis was carried out on 0.5% neutral agarose slab gel in TAE buffer (pH 8.1) at a constant current of 10mA

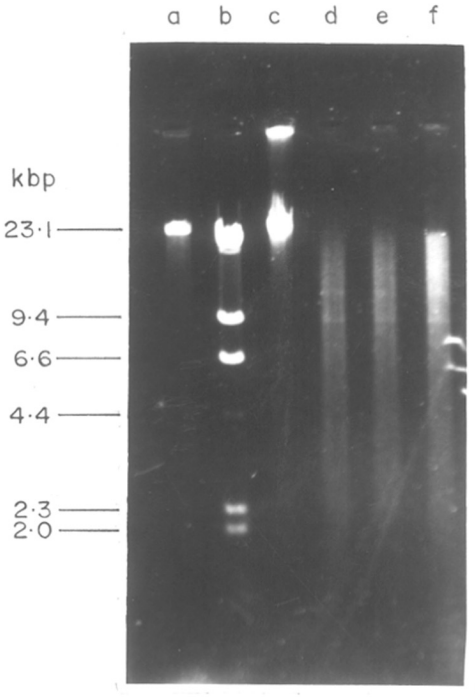


FIG. 3·1

3.1.3 Size fractionation of rice DNA on sucrose density gradient

In order to isolate DNA fragments in the size range 10-20 kbp, the partially digested rice DNA was subjected to ultracentrifugation on a linear sucrose density gradient of 10-40 % (w/w) as described in methods and materials. Fractions of 250 ul volume were collected at the end of the run. An aliquot of DNA from every third fraction was ethanol precipitated and dissolved in 10 ul TE and was analysed by electrophoresing in 0.5% agarose gel (Fig.3.2B). The DNA fractions between 6-18 were pooled, dialysed and ethanol precipitated. Before using this size fraction of EcoRI digested DNA for ligation with Charon40 arms, its ligatibility was checked (Fig.3.3A). About 80% of the fragments showed ligation indicating that the cohesive ends of rice DNA fragments were intact.

3.1.4 Preparation of vector DNA

To measure the titer , bacteriophage Charon40 lysate was titrated using E.coli LE392 host. The lysates of the titer above 10^{10} PFU/ml were used for precipitation of phages by PEG/NaCl.

As described in Methods and Materials, the phage precipitate was extracted with chloroform to remove PEG

FIGURE 3.2 A : Partial digestion of rice DNA with EcoRI

Lane a, Lane j : Lambda HindIII digest as MW marker

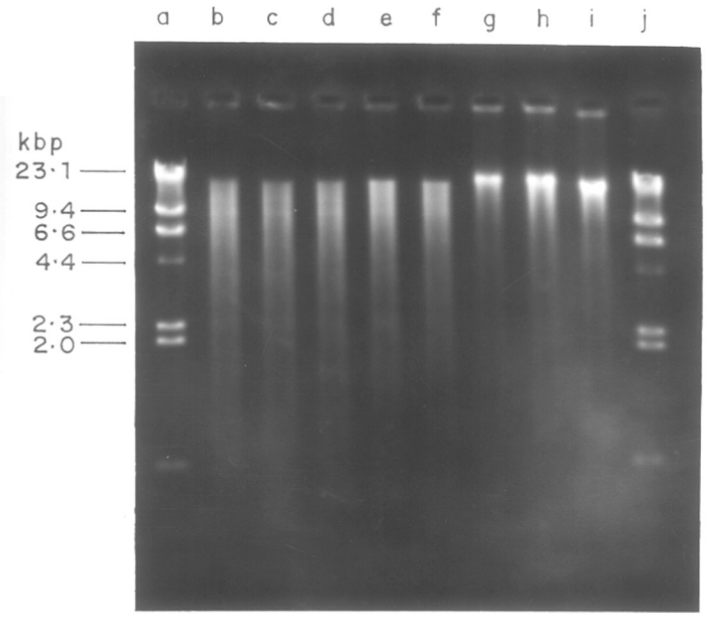
Lane b - Lane i : Rice DNA digested with decreasing units of EcoRI

FIGURE 3.2 B : Size fractionation of partially digested rice DNA on sucrose density gradient

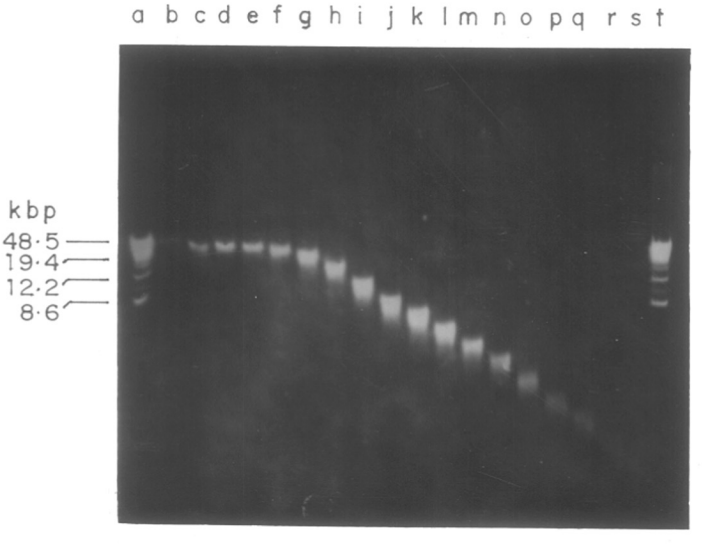
Lane a, Lane t : High molecular weight marker DNA

Lane b -Lane s : DNA from every third fraction of sucrose density gradient

Electrophoresis was carried out on 0.5% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 10mA


a b c d e f g h i j
kbp
23·1
9·4
6·6
4·4
2·3
2·0

A


a b c d e f g h i j k l m n o p q r s t
kbp
48·5
19·4
12·2
8·6

B

FIG. 3·2

FIGURE 3.3 : **Checking ligatibility of cohesive ends of rice DNA and vector DNA**

A) Lane a : High molecular weight marker DNA
Lane b : Self ligated purified rice DNA
Lane c : 10-20 kbp purified size fraction of rice DNA

B) Lane a : 1 kbp ladder as MW marker
Lane b : Charon40 arms ligated with rice DNA

Electrophoresis was carried out on 0.5% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 10mA

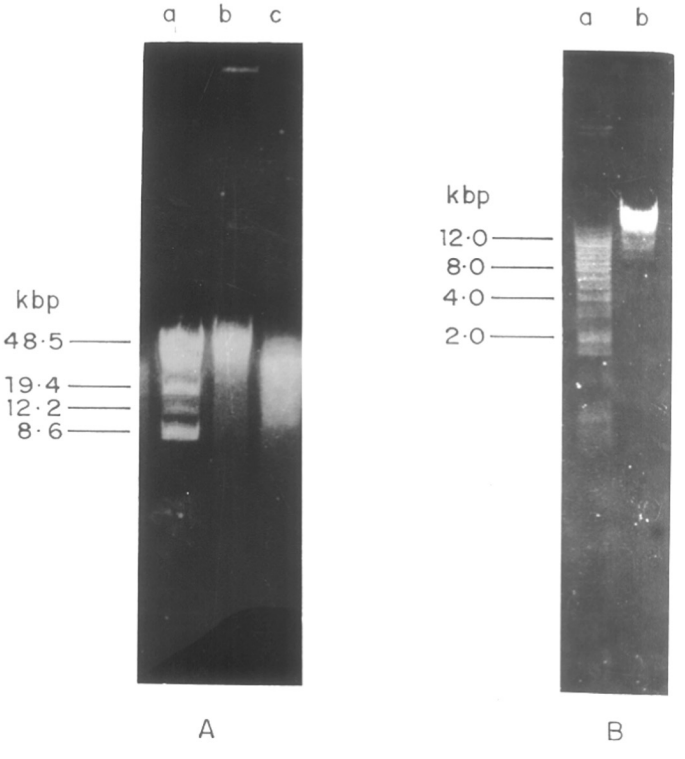


FIG. 3·3

and the concentrated phages were loaded on cesium chloride gradient for further purification. The DNA was isolated from the purified phage and its total amount and quality were determined spectrophotometrically. About 1mg DNA was obtained from 1000 ml of lysate. This DNA was free of salts and proteins. As seen in its electrophoretic pattern in 0.5% agarose gel, it did not have any RNA contamination (Fig.3.4A).

3.1.5 Preparation of vector arms

1 ug of Charon40 DNA was digested to completion with NaeI and EcoRI yielding arms and the stuffer fragments. From Fig.3.4 B (lane b) it can be seen that the Charon40 phage DNA consists of annealed left and right arms (28.8 kbp) with a stuffer fragment (235 bp). For cloning a foreign DNA, it is necessary to remove the stuffer fragment and use only the arms. EcoRI cuts Charon40 DNA into 3 fragments namely 2 arms and a stuffer fragment of molecular weight 18.8 kbp. NaeI cuts 18.8 kbp stuffer fragment into 80 smaller pieces of molecular weight 235 bp. Subsequently 100 ug of Charon40 DNA digested with NaeI and EcoRI was subjected to PEG precipitation for separation of arms from the stuffer after annealing of cos ends. Difference in the molecular weight of annealed arms and stuffers is large enough to selectively precipitate arms, leaving smaller

FIGURE 3.4 : Isolation of Charon40 DNA and its arms

A)
- Lane a : Undigested lambda DNA
- Lane b : Lambda HindIII digest
- Lane c : Charon40 DNA

B)
- Lane a : 1 kbp ladder as MW marker
- Lane b : Charon40 DNA digested with EcoRI and NaeI
- Lane c : Digested Charon40 DNA after PEG precipitation
- Lane d : High molecular weight marker DNA

C)
- Lane a : Lambda HindIII digest
- Lane b : Purified Charon40 arms

Electrophoresis was carried out on 0.5% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 10mA

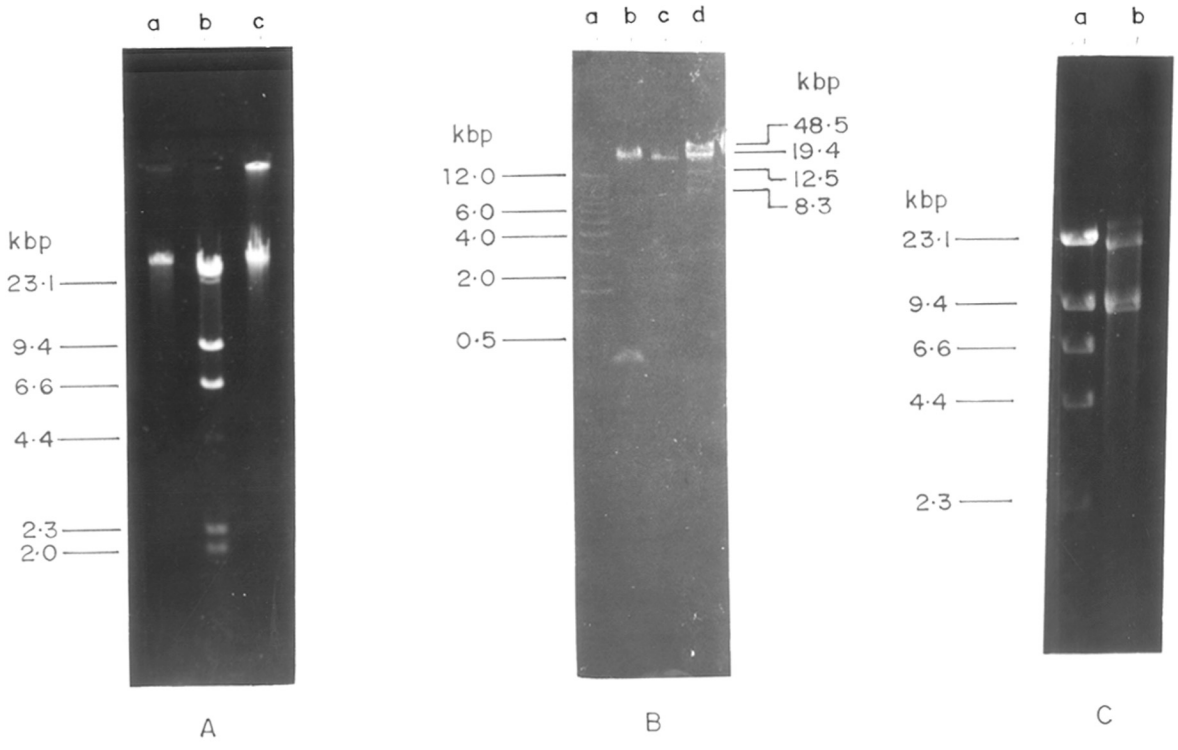

a b c
kbp
23·1
9·4
6·6
4·4
2·3
2·0
A
a b c d
kbp
12·0
6·0
4·0
2·0
0·5
kbp
48·5
19·4
12·5
8·3
B
a b
kbp
23·1
9·4
6·6
4·4
2·3
C

FIG. 3·4

stuffer fragments in solution (Fig.3.4 C ,lane c) The concentration of purified arms was determined spectrophotometrically which was 7.55 ug/20 ul. To check the purity of arms, 1 ug DNA was heated at 68°C for 10 minutes to denature cos ends and was loaded on 0.5% agarose gel along with the marker DNA. In Fig.3.4B, lane b shows annealed arms and the stuffer fragments separated on EcoRI/NaeI digestion. Figure 3.4 C (lane b) shows purified Charon 40 arms after PEG precipitation.

3.1.6 Ligation of the phage arms and the rice insert DNA

10-20 kbp size fraction of partially digested rice DNA was ligated with preannealed Charon40 arms in different vector to insert ratios i.e. 1 : 0.4, 1 : 0.6, 1 : 0.8, 1 : 1.2 (vector:insert) and the ligation was checked on 0.5% agarose gel. 1 ug vector DNA with 0.6 ug of rice DNA showed a single high molecular weight band revealing the maximum ligation (Fig.3.3B lane b). This ligation mixture was further used for in vitro packaging.

3.1.7 Preparation of in vitro packaging extract

A number of commercial packaging extracts are available which are made using two lysogens which complement each other to package externally added phage

DNA. I have prepared packaging extract from a single packaging strain E.coli SMR10 (2). During standardization, it was found that the efficiency of packaging was 1 x 10^8 when wild lambda DNA was used in the reaction. This efficiency was very close to the efficiency reported for this system (1.0-3.0 x 10^8 PFU/ug of wild lambda DNA). As mentioned in methods and materials, factors such as temperature, speed and time of growing SMR10 culture were found to be very critical to get the efficiency of the packaging extract of the order of 10^8. For example it was extremely essential to shake the culture at 300 rpm at 34°C to get the desired $O.D_{600}$ (0.8) in a minimum possible time (2.5-3 hours). The efficiency of packaging was also dependent on the time between preparation of the packaging extract and the time when it was actually used. The efficiency of packaging decreased 5-10 times when 4 weeks old packaging extract was used. In packaging experiments, therefore, the extracts used were not more than one week old. A representive data of packaging of Charon40 DNA as control and the recombinant DNA are presented in Table 3.1

3.1.8 In vitro packaging of ligated DNA

In vitro packaging of recombinant molecules of lambda Charon 40 arms and rice DNA was carried out

TABLE 3.1

In vitro packaging of Charon 40 DNA and the recombinant DNA

Phage DNA	Amount used for packaging	Dilution used	Number of Plaque per 10 ul	PFU/ug DNA
Wild lambda	lug	10^{-5}	10	1.0×10^{8}
Charon40	lug	10^{-4}	20	2×10^{7}
Charon40 arms ligated	lug	10^{-3}	1	1×10^{5}
Ligated Charon40+ rice DNA	lug	10^{-3}	9	63×10^{4}
No DNA	-	No dilution 10^{-2} 10^{-4}	-	-

using the extracts which were freshly prepared. The packaged phages were plated using E. coli ED8767 host. Ligated Charon40 arms and Charon40 DNA were used as controls for packaging. The titre of the library was 6.3×10^{5} PFU/ml. Considering 1 pg as DNA content of haploid rice genome, genome representation of the library was calculated as follows:

Rice haploid DNA content : 1.0 pg (3)

1.0 pg : 1.096×10^9 bp.

Average insert size of the recombinants : 15 kbp (1.5×10^4

so 1.5×10^4 bp : 1 recombinant phage

1.096×10^9 bp : 0.730×10^5 PFU/1 haploid rice genome

Titer of the recombinant--plaques in packaged library : 6.3×10^5 PFU.

Titer of arms packaged : 1×10^5 PFU.

Exact titer of the recombinants in the library : 5.3×10^5 PFU.

0.730×10^5 PFU represents one rice genome. Therefore, 5.3×10^5 PFU represents 7.26 genomes of rice. The genomic library of rice constructed by me thus represents 7.6 genomes of rice.

SECTION II

Screening of the rice genomic library and identification of clones containing glutelin gene

Rice genomic library which represents 7.6 genomes of rice has the enormous population of recombinants. I was interested in identification and isolation of clones encoding the major seed storage protein of rice "glutelin" from this pool of recombinants. The most commonly used method for isolating the desired clones involves screening of bacteriophage plaques by hybridization with P-32- labelled probes.

3.2.1 Amplification of the library

Prior to screening, amplification of the library was done to get overrepresentation of the low copy glutelin clones to increase the probability of there identification .

Rice genomic library was amplified using E. coli ED8767 as host. The titer of the amplified library was appoximately 10^{10} PFU/ml of the lysate.

3.2.2 Use of oligonucleotide probe for screening

Different types of probes used for screening the genomic library are heterologous probes, cDNA probes or

oligonucleotide probe. I have used oligonucleotide probe since cDNA or heterologous probes were not available with us initially.

The oligoprobe was synthesized from the available cDNA sequence of glutelin gene of japonica rice (4). It was purified on 14% polyacrylamide gel. DNA bands were visualized by UV shadowing with a short wave UV lamp. The 33 bases band was the slowest running major band. This band was eluted from the gel and dissolved to get oligonucleotide probe of final concentration 1 ug/ul.

The sequence of the oligoprobe is

GAGCCAATTCGGAGTGTGAGGTCTCAAGCTGGC

with G+C content of 57.5%.

This 33 bases sequence selected to make oligoprobe was present in the 5' end of the coding region of glutelin gene. The high G+C content of the probe helped to obtain stable hybrids. The probe was labelled using P-32-dATP in a kinasing reaction and used for hybridization.

3.2.3 Primary screening and plaque purification

Identification of clones containing glutelin gene from amplified rice genomic library was done by carrying out hybridization of plaques.

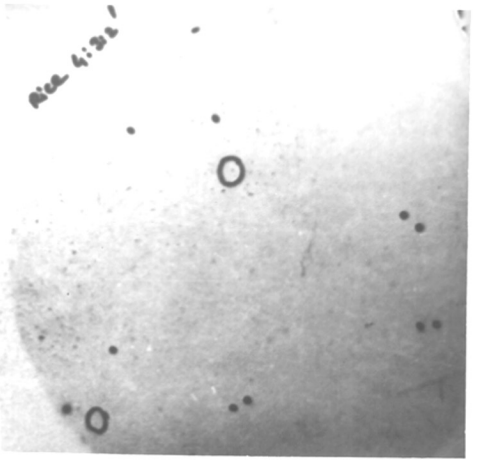
For this purpose, the amplified library was

plated at an approximate density of 10,000 plaques per 150 mm plate. Plaque lifts in duplicate from 10 such plates were made which were used for hybridization with oligoprobe. 13 hybridizing plaques identified by aligning films with original plates were picked up for further analysis. This was the preliminary round of screening where about 10^5 recombinant phages were screened (Fig.3.5A). In the second round of plaque purification,13 plaques showing positive signals in first screening were plated separately to get about 1000 plaques per plate. Duplicate plaque lifts were prepared from these plates which were hybridized with oligoprobe. Out of 13 clones, only 7 gave positive signals consistently till the last round of purification where all the 100 plaques per plate showed positive signals (Fig.3.5B). Thus seven probable clones of rice glutelin hybridizing to glutelin oligonucleotide probe were obtained after screening the library which corresponded to about 1 rice genome (about 10^5 PFU). These seven clones were named as lambda RG1, lambda RG2, lambda RG3, lambda RG4, lambda RG5, lambda RG6 and lambda RG7.

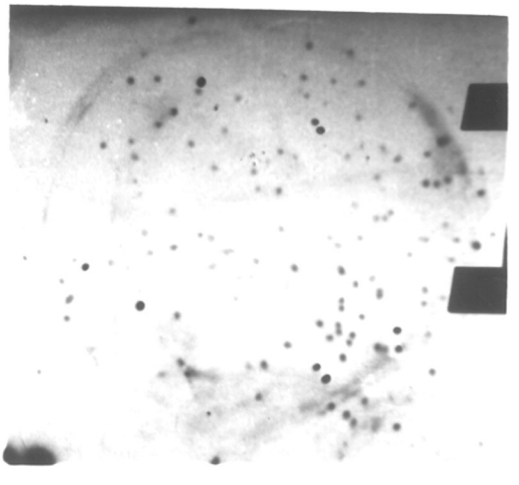
URE 3.5 : **Autoradiogram of plaque hybridization for screening of rice genomic library with oligonucleotide probe**

A) 1st round of plaque hybridization

B) Last round of plaque purification



A



B

FIG. 3·5

3.2.4 Dot blot analysis of seven clones for confirming the presence of sequence encoding glutelin

DNAs of all the seven glutelin clones were isolated and checked on 0.5% agarose gel (Fig.3.6). All of them show high molecular weight band of about 45 kbp which includes two arms and the insert of about 15 kbp. Since oligonucleotide probe is not very specific because of its smaller size, it was necessary to confirm the presence of glutelin gene in these clones. By this time rice glutelin cDNA form japonica variety (5) was kindly made available to me by Prof. Muthukrishnan, Kansas State University, Kansas, USA which I used for confirming the clones. Lambda RG1 to lambda RG7 DNAs were denatured and dot blotted along with the positive controls of rice DNA and glutelin cDNA. Negative control of Charon40 DNA was spotted on the same blot. Hybridization was carried out using glutelin cDNA insert as a probe. Out of seven glutelin clones , only three clones namely lambda RG4, lambda RG6, and lambda RG7 showed positive signals confirming the presence of glutelin gene in these clones (Fig.3.7). These three clones were used for further characterization.

FIGURE 3.6: Isolation of DNAs of seven glutelin clones namely Lambda RG1 to Lambda RG7

Lane a :	Lambda HindIII digest	Lane e :	Lambda RG4 DNA
Lane b :	Lambda RG1 DNA	Lane f :	Lambda RG5 DNA
Lane c :	Lambda RG2 DNA	Lane g :	Lambda RG6 DNA
Lane d :	Lambda RG3 DNA	Lane h :	Lambda RG7 DNA

Electrophoresis was carried out on 0.5% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 10mA

FIGURE 3.7: Dot blot hybridization of DNAs of 7 glutelin clones with glutelin cDNA probe

1	Rice DNA
2	Glutelin cDNA
3	Lambda RG4 DNA
4	Lambda RG6 DNA
5	Lambda RG7 DNA

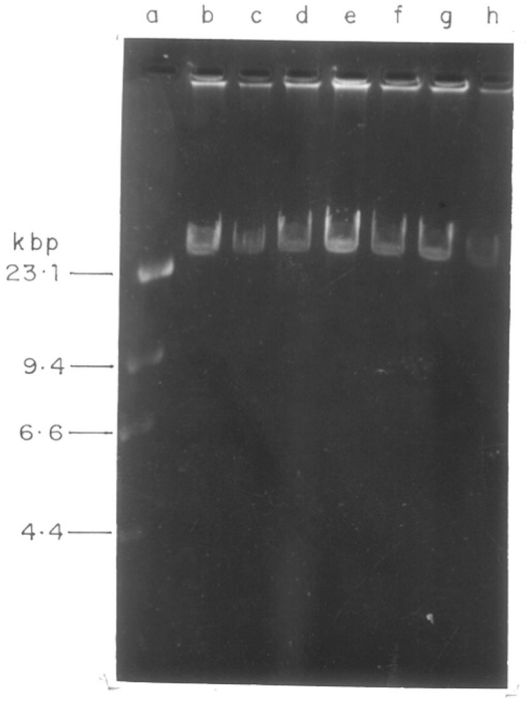


FIG. 3·6

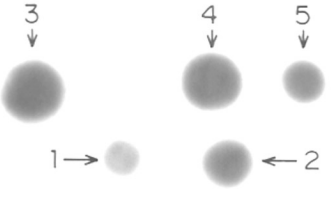


FIG. 3·7

SECTION III

Characterization of rice (Basmati-370) glutelin gene

DNAs from the seven clones of glutelin were isolated out of which 3 clones namely lambda RG4, lambda RG6 and lambda RG7 gave strong signals on hybridization with glutelin cDNA from japonica rice (5). One of these three clones namely lambda RG4 was selected for its characterization with respect to restriction mapping, copy number estimation and preliminary nucleotide sequencing.

3.3.1 Determination of insert sizes in 3 clones lambda RG4, lambda RG6, and lambda RG7

The presence and sizes of inserts in the 3 clones were determined by digesting lambda RG4, lambda RG6 and lambda RG7 DNAs with EcoRI to separate arms from the inserts From Fig.3.8, it is seen that lambda RG6 and lambda RG7 give similar band pattern with two bands of 19.2 kbp and 9.6 kbp corresponding to left and right arms respectively and 2 more bands of 9.0 kbp and 1.55 kbp showing presence of one EcoRI site in the inserts. Lambda RG4 shows presence of left and right arms and two insert bands of 9.4 kbp and 3.55 kbp. All the three clones thus have inserts of sizes around 10-13 kbp with

FIGURE 3.8 : Determination of insert sizes in 3 glutelin clones

A)
Lane a : Lambda HindIII digest
Lane b : EcoRI digested Lambda RG4 DNA

B)
Lane a : 1 kbp ladder as MW marker
Lane b : Lambda RG6 DNA
Lane c : EcoRI digested Lambda RG6 DNA
Lane d : Lambda RG7 DNA
Lane e : EcoRI digested Lambda RG7 DNA

Electrophoresis was carried out on 0.5% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 10mA

Arrows indicate the DNA fragments from inserts

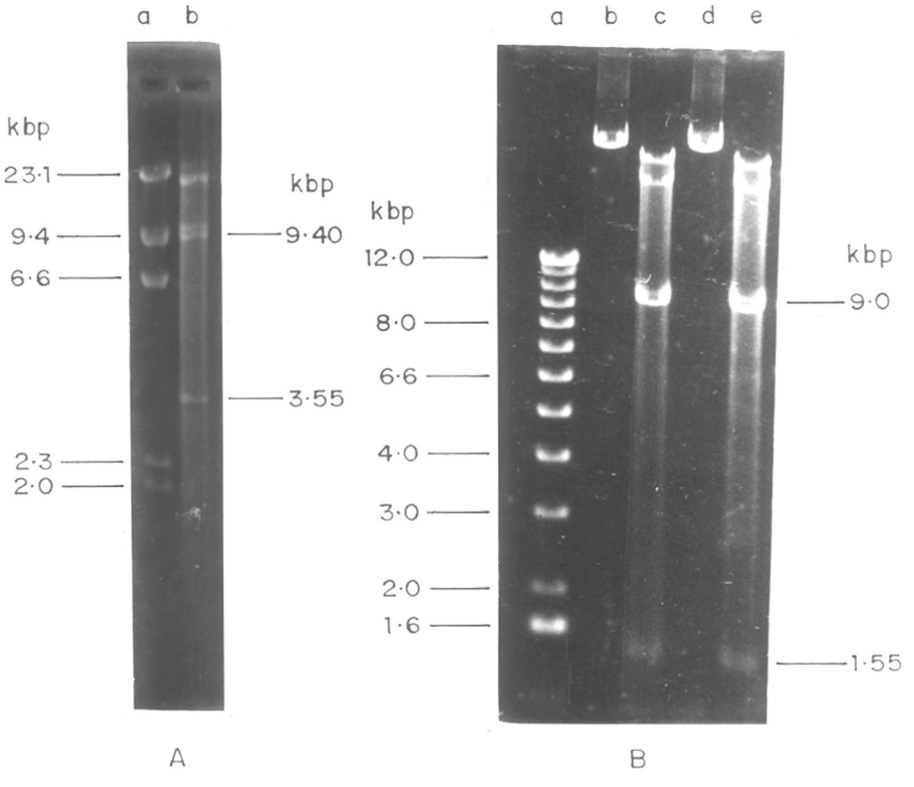

a
b
kbp
23·1
9·4
6·6
2·3
2·0
kbp
9·40
3·55
A
a
b
c
d
e
kbp
12·0
8·0
6·6
4·0
3·0
2·0
1·6
kbp
9·0
1·55
B

FIG. 3·8

one EcoRI site.

3.3.2 Restriction enzyme analysis of lambda RG4

Restriction enzymes from the multiple cloning sites region of the vector Charon40 were used for the digestion of lambda RG4 DNA. EcoRI was always used along with second restriction endonuclease to separate out arms. Lambda RG4 DNA was digested with EcoRI, EcoRI/HindIII, EcoRI/KpnI, EcoRI/BamHl, EcoRI/SmaI, and EcoRI/XbaI. Figure 3.9 shows the electrophoresis pattern of lambda RG4 DNA digested with each of these enzymes. In case of EcoRI and BamHI double digestion (lane e),the pattern is similar to EcoRI digestion (lane b) showing absence of BamHI site in the inserts of lambda RG4. In EcoRI/HindIII, EcoRI/KpnI, EcoRI/SmaI, and EcoRI/XbaI digests 9.4 kbp as well as 3.55 kbp bands disappear and many additional smaller bands are seen indicating the presence of many sites for these enzymes in the two inserts.

3.3.3 Identification of the fragment containing glutelin gene from lambda RG4

As stated earlier, out of the three clones namely

FIGURE 3.9 : **Restriction endonuclease analysis of Lambda RG4**

Lane a : Lambda HindIII digest
Lane b : Digestion with EcoRI
Lane c : Digestion with EcoRI/HindIII
Lane d : Digestion with EcoRI/KpnI
Lane e : Digestion with EcoRI/BamHI
Lane f : Digestion with EcoRI/SmaI
Lane g : Digestion with EcoRI/XbaI

Electrophoresis was carried out on 0.5% neutral agarose slab gel in TAE buffer (pH 8.1) at a constant current of 10mA

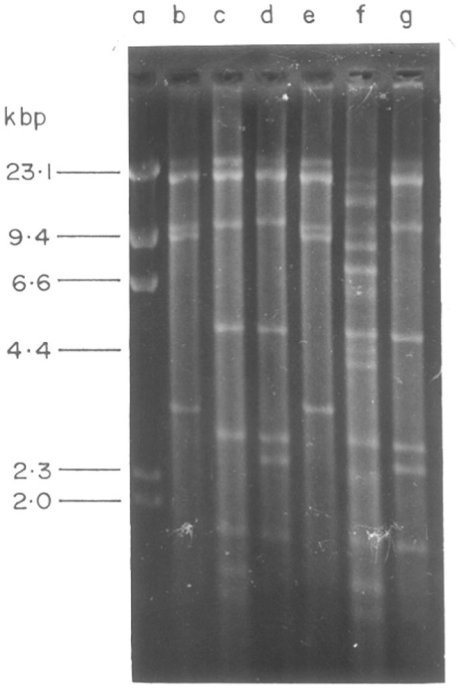


FIG. 3·9

lambda RG4, lambda RG6 and lambda RG7; lambda RG4 was selected randomly to carry out further characterization.

Lambda RG4 DNA when digested with EcoRI gives two insert bands. To determine which of the 2 bands contained the glutelin gene, hybridization with glutelin cDNA from japonica rice (5) was carried out. It is clearly seen from Fig. 3.10B (lane b) that only one band of 3.55 kbp gives the positive signal revealing the presence of glutelin gene in it. Since a single band of 3.55 kbp shows the hybridization with glutelin cDNA, it indicates the intactness of the gene.

3.3.4 Subcloning of 3.55 kbp EcoRI fragment in plasmid Bluescript

The 3.55 kbp EcoRI fragment containing the glutelin gene was subcloned in pBS KS^+ for further characterization . The plasmid bluescript KS^+ was selected for subcloning because of easy screening procedures and direct use of the cloned fragment for expression studies and sequencing (6).

For subcloning, the 3.55 kbp band was eluted from the gel (Fig 3.11A) and ligated to the EcoRI restricted, dephosphorylated vector DNA (Fig.3.11B). Figure 3.11 C shows the occurrence of ligation.

FIGURE 3.10 A: Electrophoresis pattern of Lambda RG4 DNA digested with EcoRI

Lane a : Lambda HindIII digest
Lane b : Lambda RG4 DNA digested with EcoRI

Electrophoresis was carried out on 0.5% neutral agarose slab gel in TAE buffer (pH 8.1) at a constant current of 10mA

FIGURE 3.10 B : Southern hybridization of P-32 labelled glutelin cDNA with Lambda RG4 DNA digested with EcoRI

Lane wise order is same as described in Fig.3.10 A

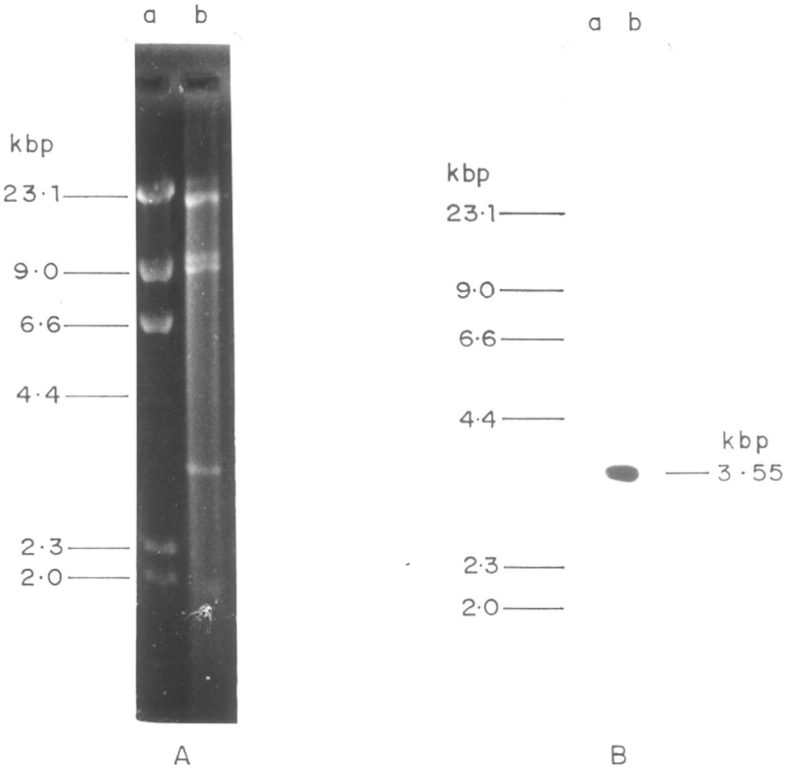


FIG. 3·10

FIGURE 3.11 : **Subcloning of 3.55 kbp fragment containing glutelin gene**

A) Lane a : Lambda HindIII digest as MW marker
Lane b : Eluted 3.55 kbp fragment

B) Lane a : 1 kbp ladder as MW marker
Lane b : Plasmid vector Bluescript KS^+
Lane c : pBS KS^+ digested with EcoRI
Lane d : pBS KS^+ EcoRI digested and dephosphorylated

C) Lane a : Lambda HindIII digest as MW marker
Lane b : Ligated pBS KS^+ and 3.55 kbp fragment

Electrophoresis was carried out on 0.7% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 30mA

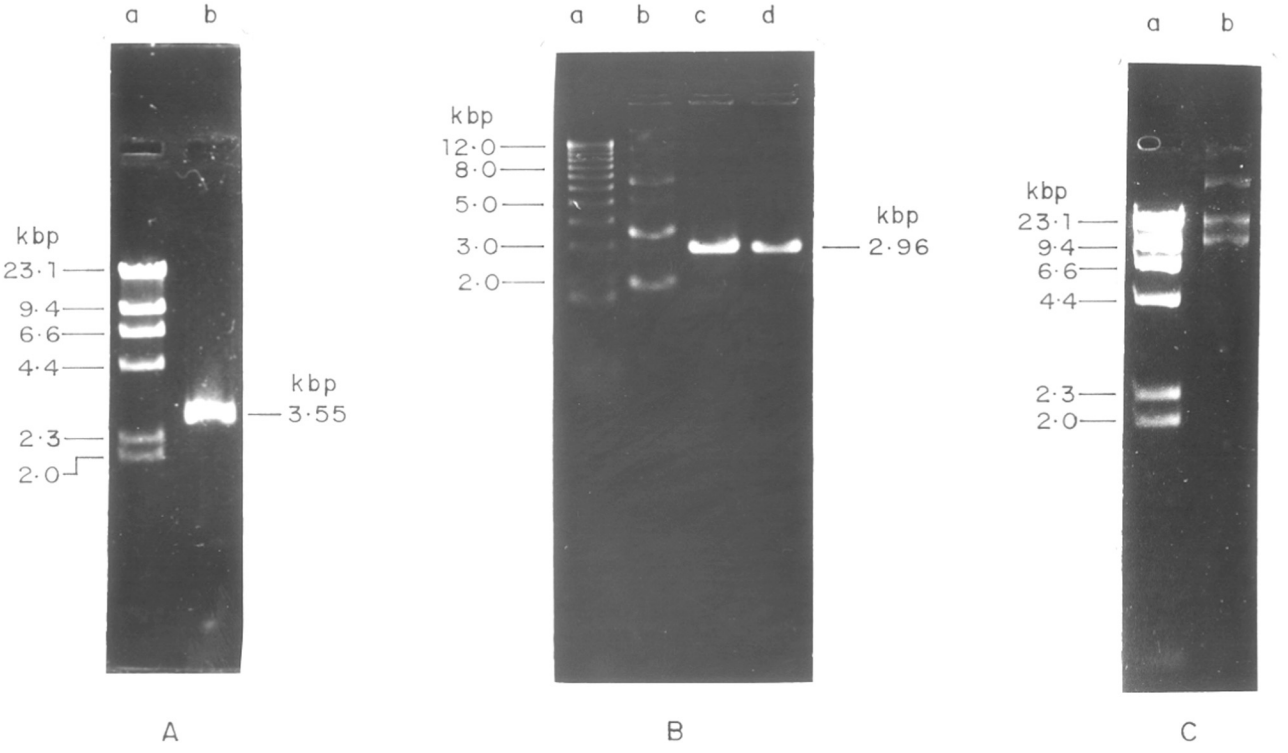

a
b
kbp
23·1
9·4
6·6
4·4
2·3
2·0
kbp
3·55
A
a
b
c
d
kbp
12·0
8·0
5·0
3·0
2·0
kbp
2·96
B
a
b
kbp
23·1
9·4
6·6
4·4
2·3
2·0
C

FIG. 3·11

Chimeric plasmids were transformed into E.coli DH5α and the recombinants were selected on X-gal-ampicillin LB plates. The plasmid containing 3.55 kbp insert was named as pRG4.

3.3.5 Restriction mapping of pRG4

In order to confirm the presence of insert in the recombinant plasmid, DNA from the recombinant clone was isolated and digested with EcoRI followed by electrophoresis on 0.7% agarose gel (Fig.3.12A). The DNA band at a fragment length of 2.96 kbp represents the linearised bluescript and 3.55 kbp band represents the insert.

Different hexacutter restriction endonucleases were used for mapping the pRG4 sequence. Initial mapping was done using the enzymes which have unique sites and are from the multiple cloning sites region of the vector pBS. Figure 3.12B shows the electrophoresis pattern of pRG4 sequence digested with XbaI, SalI, PstI, NotI, BamHI and ClaI. Here a single fragment of 6.5 kbp is seen, i.e. there is just linearization of the cloned sequence, thus indicating the absence of sites for these restriction enzymes in the pRG4 sequence. Restriction endonucleases like BclI, BglII, BstEII and NcoI which have no sites in bluescript were next used for digestion of pRG4. These

FIGURE 3.12 : **Digestion of pRG4 with different restriction endonucleases**

A)
Lane a : 1 kbp ladder as MW marker
Lane b : pRG4 digested with EcoRI

B)
Lane a : 1 kbp ladder as MW marker
Lane b : XbaI
Lane c : SalI
Lane d : PstI
Lane e : NotI
Lane f : BamHI
Lane g : ClaI

C)
Lane a : 1 kbp ladder as MW marker
Lane b : BclI
Lane c : BglII
Lane d : BstEII
Lane e : NcoI

Electrophoresis was carried out on 0.7% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 30mA

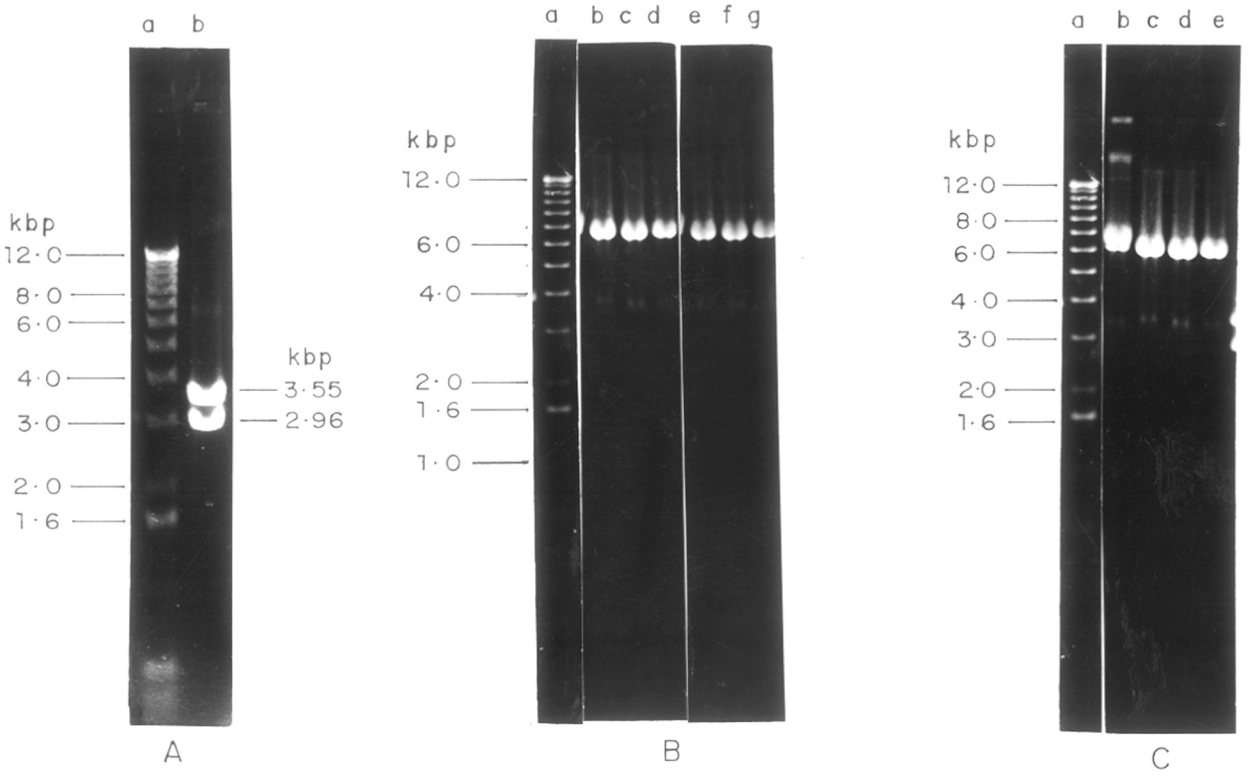

a b
kbp
12·0
8·0
6·0
4·0
3·0
2·0
1·6
kbp
3·55
2·96
A
a b c d e f g
kbp
12·0
6·0
4·0
2·0
1·6
1·0
B
a b c d e
kbp
12·0
8·0
6·0
4·0
3·0
2·0
1·6
C

FIG. 3·12

digests show 6.5 kbp linearised band as seen in Fig.3.12 C suggesting the presence of one site for each of these restriction endonucleases in pRG4 insert.

Figure 3.13 depicts the digestion patterns of pRG4 DNA with AvaI, BglI , ScaI , EcoRV , PvuII and PvuI In the vector pBS there are two sites at 668 bp and 713 bp for AvaI and 472 bp and 2166 bp for BglI. Figure 3.13 A lane b shows 3 bands of 2.98 kbp, 2.0 kbp and 1.54 kbp in AvaI digest and 3 fragments of 3.23 kbp,1.97 kbp and 1.31 kbp in BglI digest (lane c) indicating the presence of one site for AvaI and BglI in the insert. ScaI digestion shows two bands of sizes 4.25 kbp and 2.25 kbp revealing one ScaI site in the insert (Fig. 3.13 B).

Among other restriction enzymes tried for mapping were enzymes like EcoRV (1 site in pBS at 697 bp) PvuI (cuts pBS at 500 bp and 2416 bp) and PvuII (cuts pBS at 529 bp). PvuI and PvuII digests show many bands suggesting more than two sites in the insert (Fig. 3.13 D,lanes b and c) . In EcoRV digest (Fig 3.13 C), there are 3 bands of 4.04 kbp, 2.0 kbp and 0.46 kbp indicating presence of two EcoRV sites in the sequence.

Restriction digestions of pRG4 with respect to XhoI, BstEII, BglI and EcoRV are depicted in Figs.3.14

FIGURE 3.13 : Digestions of pRG4 DNA with different restriction endonucleases

A) Lane a : 1 kbp ladder as MW marker
Lane b : AvaI
Lane c : BglI

B) Lane a : 1 kbp ladder as MW marker
Lane b : ScaI

C) Lane a : 1 kbp ladder as MW marker
Lane b : EcoRV

D) Lane a : 1 kbp ladder as MW marker
Lane b : PvuII
Lane c : PvuI

Electrophoresis was carried out on 0.7% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 30mA

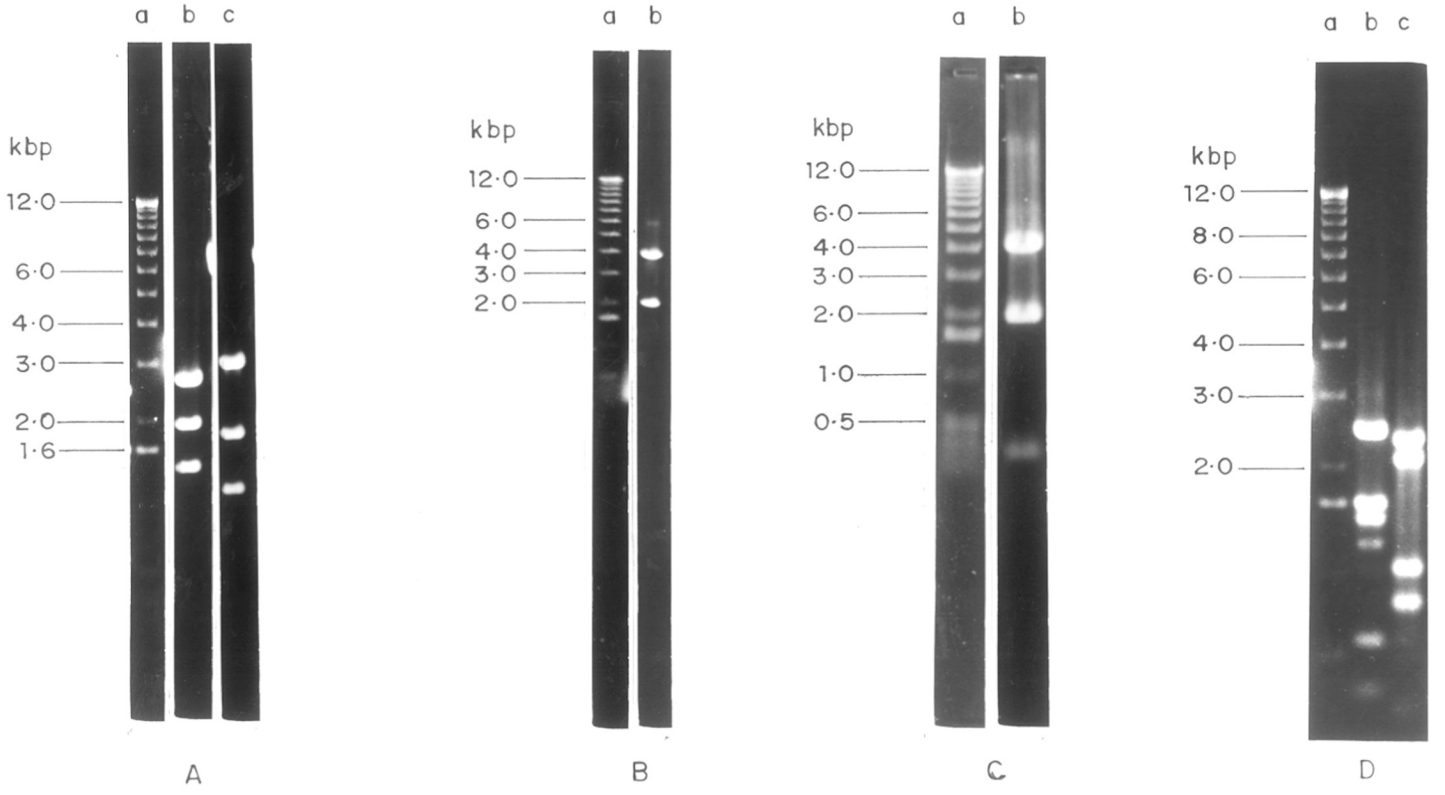

a b c
kbp
12·0
6·0
4·0
3·0
2·0
1·6
A
a b
kbp
12·0
6·0
4·0
3·0
2·0
B
a b
kbp
12·0
6·0
4·0
3·0
2·0
1·0
0·5
C
a b c
kbp
12·0
8·0
6·0
4·0
3·0
2·0
D

FIG. 3·13

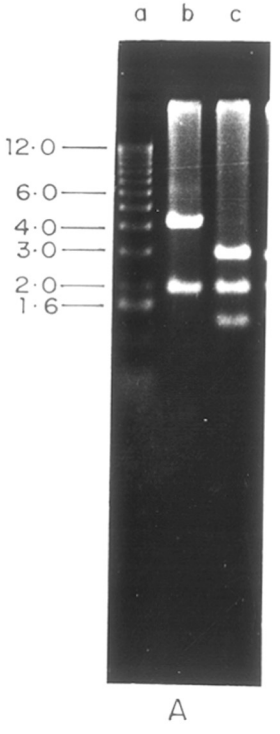
FIGURE 3.14 : **Restriction enzyme digestions of pRG4**

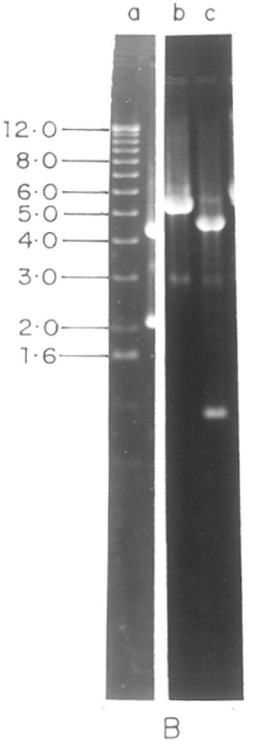
A) Lane a : 1 kbp ladder as MW marker
Lane b : XhoI
Lane c : XhoI/EcoRI

B) Lane a : 1 kbp ladder as MW marker
Lane b : Bst EII
Lane c : Bst EII/XbaI

C) Lane a : 1 kbp ladder as MW marker
Lane b : Bst EII/EcoRI

Electrophoresis was carried out on 0.7% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 30mA


a
b
c
12·0
6·0
4·0
3·0
2·0
1·6
A


a
b
c
12·0
8·0
6·0
5·0
4·0
3·0
2·0
1·6
B


a
b
12·0
8·0
5·0
3·0
2·0
1·6
C

FIG. 3·14

and 3.15. XhoI digests pRG4 to give two bands of 4.48 kbp and 2.0 kbp because of 1 XhoI site in pBS at 668 bp and another site in pRG4 insert. EcoRI/XhoI double digestion shows 3 bands of molecular weight 2.96 kbp, 1.97 kbp and 1.51 kbp where former corresponds to linearised vector while the latter two are inserts. As seen in XhoI digest, high molecular weight band (4.48 kbp) could appear only if XhoI site in the insert is at 1.97 kbp from XhoI site at 668 bp in pBS. XhoI site is mapped with respect to XhoI site in pBS (Fig 3.16). Restriction enzyme Bst EII has no site in pBS vector. The absence of Bst EII in the insert of pRG4 would then give a picture with no digestion. The digestion, however, shows linearization of the plasmid with a single band at 6.5 kbp suggesting the presence of Bst EII site in the cloned insert (Fig.3.14B) (lane b).

Figures 3.14 B (lane c) and C (lane b) depict the double digestion patterns of pRG4 with BstEII/XbaI and BstEII/EcoRI respectively. In BstEII/XbaI double digestion 2 bands of 5.37 kbp and 1.21 kbp are obtained. Here the 5.37 kbp band could appear only if BstEII site is at 2.27 kbp from the EcoRI site. In BstEII/EcoRI double digestion 3 bands of 2.96 kbp, 2.41 kbp and 1.21 kbp are seen. 2.96 kbp is linearised pBS while 2.41 and 1.21 kbp bands are because of the single BstEII site in the insert.

Figures 3.15 A and B include the restriction enzyme digestion patterns of pRG4 with BglI, EcoRI/BglI, EcoRV and EcoRI/EcoRV. The plasmid bluescript has 2 BglI sites at 472 bp and 2166 bp. Since insert has one Bgl site in it, BglI digestion of pRG4 gives 3 bands of 3.25 kbp, 1,97 kbp and 1.26 kbp (Fig.3.15A) (lane b). Double digestion of pRG4 with EcoRI/BglI gives 3 bands of 279 bp, 1.25 kbp and 1.48 kbp because of two BglI sites in the pBS and two bands of 2.9 kbp and 639 bp because of one BglI site in the insert which is mapped as shown in Fig.3.16.

In EcoRV single digest of pRG4, 3 bands of 4.11 kbp, 1.95 kbp and 435 bp are obtained (lane b). In EcoRI/EcoRV double digestion, 5 bands with 1.95 kbp and 435 bp bands common in both the digests are seen. Moreover, 1.18 kbp band along with 2.96 kbp plasmid band together form the 4.11 kbp band which is seen in the single digest. Finally, one EcoRV site is mapped at 2389 bp from EcoRI end as shown in Fig. 3.16.

Figure 3.15 C shows double digestion of pRG4 with EcoRI/HindIII (lane b); EcoRI/KpnI (lane d) and EcoRI/SmaI (lane f). In all the 3 digests, 3 bands of 2.98 kbp, 2.0 kbp and 1.55 kbp are seen indicating presence of one site for each of these enzymes in pRG4 insert.

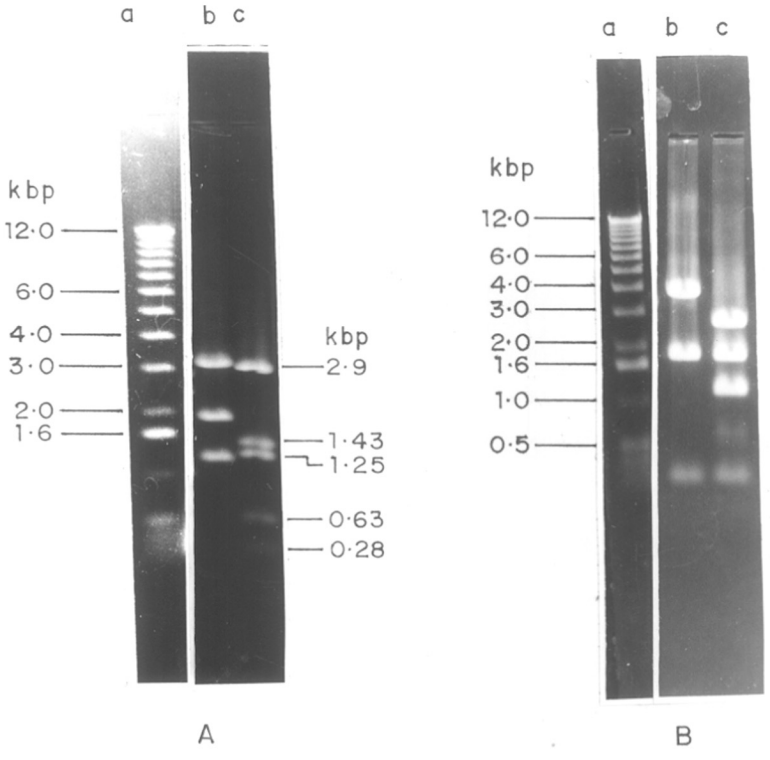
FIGURE 3.15 Restriction enzyme digestions of pRG4

A) Lane a : 1 kbp ladder as MW marker
Lane b : BglI
Lane c : BglI/EcoRI

B) Lane a : 1 kbp ladder as MW marker
Lane b : EcoRV
Lane c : EcORI/EcoRV

C) Lanes a,c,e : 1 kbp ladder as MW marker
Lane b : EcoRI/HindIII
Lane d : EcoRI/KpnI
Lane f : EcoRI/SmaI

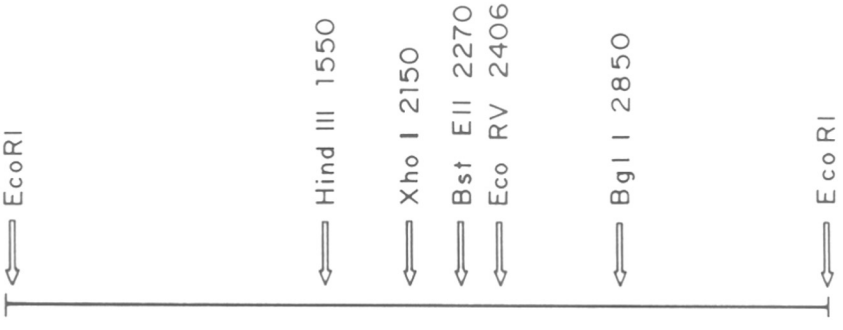
Electrophoresis was carried out on 0.7% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 30mA


a
b c
kbp
12·0
6·0
4·0
3·0
2·0
1·6
kbp
2·9
1·43
1·25
0·63
0·28
A
a
b
c
kbp
12·0
6·0
4·0
3·0
2·0
1·6
1·0
0·5
B


a b c d e f
kbp
12·0
6·0
3·0
2·0
1·6
C

FIG. 3·15

FIGURE 3.16 Restriction endonuclease map of pRG4



SCALE : 1 kbp = 3·75 cms

FIG. 3·16

Based on these studies, it can be concluded that 3.55 kbp insert of pRG4 contains no sites for XbaI, SalI, PstI, NotI, BamH1; one site for AvaI, KpnI, HindIII , ClaI, BclI, BglII, BstEII, NcoI, XhoI, BglI, and SmaI and 3 sites for EcoRV, PvuI and PvuII.

3.3.6 Genomic reconstruction analysis of pRG4

Southern hybridization was carried out to investigate the genomic organization of the glutelin gene and to estimate its copy number. Rice leaf DNA was digested with EcoRI, HindIII, BamH1 and HimfI (Fig.3.17A) and hybridized with P-32-labelled pRG4 insert DNA (Fig.3.17B). The pRG4 insert hybridizes with 2 EcoRI fragments (3.55 kbp and 1.5 kbp), 3 HindIII fragments (9.0 kbp, 7.0 kbp, 3.55 kbp) and 1 BamH1 fragment (1.5 kbp). The occurrence of 2-3 hybridizing bands depicts the nontandem or dispersed nature of this gene (pRG4) and further indicates that there are two or more glutelin gene subfamilies with totally different environment in the rice genome. The presence of 2 or 3 subfamilies differing in size and sequence has been reported in rice (4). The 3.55 kbp band in EcoRI digest is due to the glutelin gene family represented by pRG4 clone. The 1.55 kbp band in HindIII digest especially indicates the presence of HindIII site in

the 3.55 kbp insert which is confirmed in restriction mapping . The 1.5 kbp band in EcoRI digest is probably due to the family which is different from pRG4 family because 3.55 kbp does not have internal EcoRI site. BamHl digest shows single band of 1.5 kbp which appeared probably because of internal BamHI site in 3.55 kbp glutelin gene. However , I have not observed BamHI site in the pRG4 insert. Therefore, it is possible that this band represents second family of glutelin gene which has homology with pRG4 . In densitometric scans, when the total area under the peaks in the genomic lanes is compared to the area in the reconstruction lanes, a rough estimate of the total of 5-7 copies of glutelin gene per haploid rice genome can be made (Fig.3.18). When genomic blot was probed with glutelin cDNA exactly same band pattern was observed suggesting the unique organization of 5' and 3' flanking sequences which are present exclusively along with the glutelin genes in rice genome.

FIGURE 3.17 : Copy number estimation of glutelin gene

A)

Lane a : 1 kbp ladder as MW marker
Lane b : EcoRI digest of rice DNA
Lane c : HindIII digest of rice DNA
Lane d : BamHI digest of rice DNA
Lane e : HinfI digest of rice DNA
Lane f : 1 copy of 3.55 kbp pRG4 insert
Lane g : 2 copies of 3.55 kbp pRG4 insert
Lane h : 5 copies of 3.55 kbp pRG4 insert
Lane i : 10 copies of 3.55 kbp pRG4 insert
Lane j : 20 copies of 3.55 kbp pRG4 insert

Electrophoresis was carried out on 0.7% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 30mA

B) Southern hybridization of 32-P-labelled pRG4 insert with rice DNA digested with different restriction endonucleases

Lane wise order is same as described in Fig. 3.17 A.

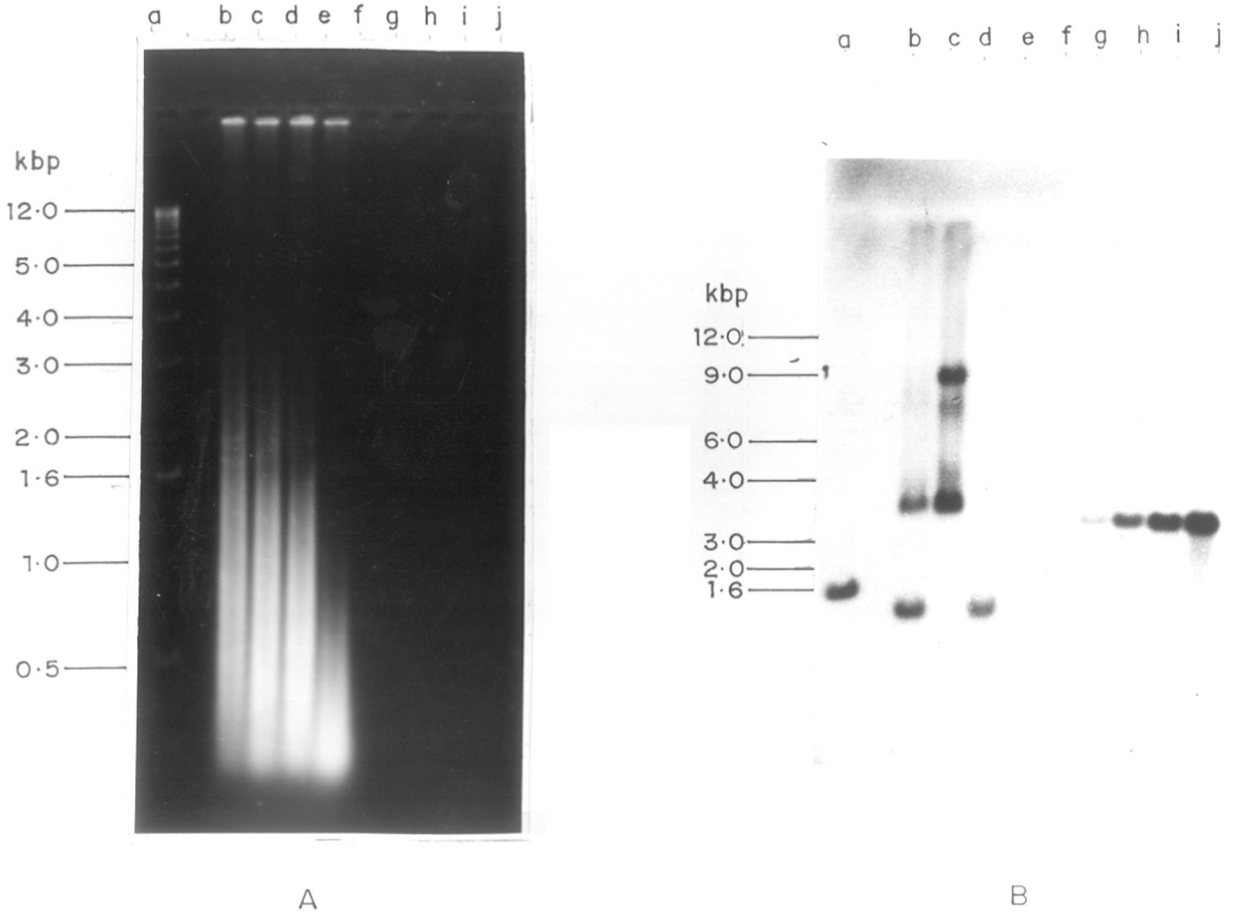

a b c d e f g h i j
kbp
12·0
5·0
4·0
3·0
2·0
1·6
1·0
0·5
A
a b c d e f g h i j
kbp
12·0
9·0
6·0
4·0
3·0
2·0
1·6
B

FIG. 3·17

FIGURE 3.18 : Densitometric scans of the autoradiogram shown in Fig. 3.17 B

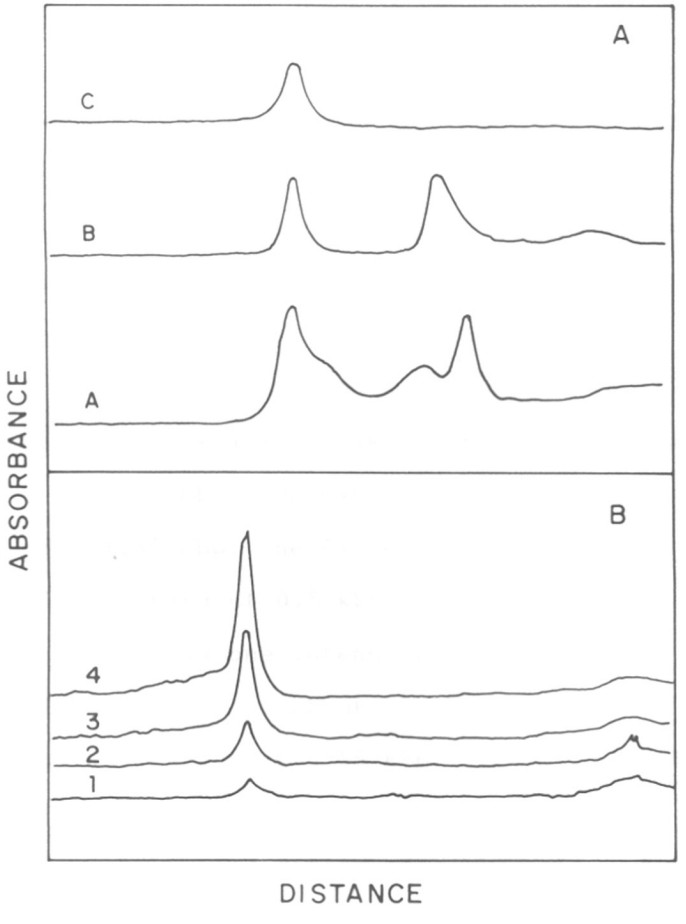


FIG. 3·18

3.3.7 Presence and distribution of pRG4 sequence in two subspecies of rice namely Basmati 370 (indica) and Taichung 65 (japonica)

Uptil now glutelin genes from japonica varieties of rice have been cloned and characterized. I have cloned the glutelin gene from indica variety of rice basmati 370. I wanted to know if there is any difference in the organization of this particular glutelin gene in indica and japonica varieties. Rice DNAs of varieties Taichung 65 (japonica) and Basmati 370 (indica) were digested with EcoRI (Fig.3.19A) and southern hybridization was carried out using pRG4 insert as probe . Basmati 370 DNA shows one strong band of 3.55 kbp, one faint band of 1.5 kbp and one very faint band of 0.5 kbp. Taichung 65 DNA, on the other hand, shows one intense band of 3.55 kbp and one very faint band of about 5 kbp (Fig.3.19B). This pattern indicates that 3.55 kbp family of glutelin has probably the same organization in Basmati 370 and Taichung 65. Second family of glutelin represented by 1.5 kbp band is either absent in Taichung 65 or has different environment in Taichung 65 genome giving high molecular weight band of 5 kbp. It is necessary to find out if there is any difference in glutelin genes of indica and japonica rices. This could be established

FIGURE 3.19 A : **Digestion of DNAs of Basmati 370 (indica) and Taichung 65 (japonica) varieties of rice**

Lane a : 1 kbp ladder as molecular weight marker
Lane b : Basmati 370 DNA digested with EcoRI
Lane c : Taichung 65 DNA digested with EcoRI

Electrophoresis was carried out on 0.7% neutral agarose slab gels in TAE buffer (pH 8.1) at a constant current of 30mA

FIGURE 3.19 B : **Autoradiogram showing the hybridization of pRG4 insert with rice Basmati-370 DNA and Taichung 65 DNA digested with EcoRI**

Lane wise order is same as described in Fig.3.19 A

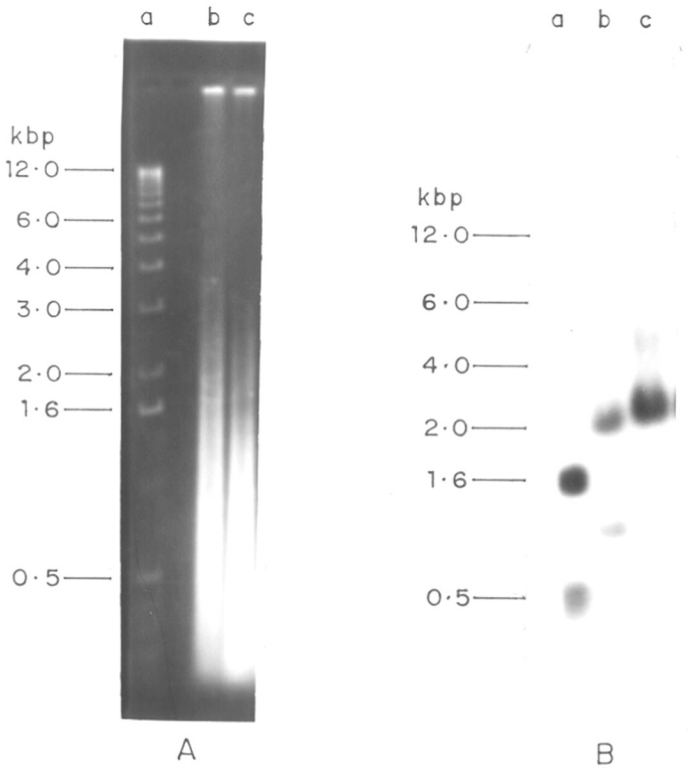


FIG. 3·19

only after sequencing the glutelin gene.

3.3.8 DNA sequencing of pRG4

To study the mechanisms of regulation of genes encoding glutelins, it is necessary to determine the primary nucleotide sequence of coding as well as noncoding regions within the gene and also of the 5' upstream regions containing the regulatory elements.

I have sequenced about 200 bp of pRG4 by Sangers dideoxy nucleotide chain termination method as described earlier in Chapter 2. The sequence obtained from 5' ---> 3' end is as follows :

Towards 5' end of the insert.

```
5'_ _ _ _ _ C C C C G G G C T A C C C G C C T C G C
          A A C T T C C T C C T C C T T A C T C T G C
          A A A A C C G T A A A G T G G T T G C C T G
          G G C C T A A A C A G C A C G G C A T C C A
          T A T T A C T T C C T A T A T G A C T G C G
          C G C T A T G G T A A G G C T G A G A T G A
          G G A C T T C G C T G T A T C G A G C T A A
          C A A T C G G A C T C G A C A G T C G A T C
          T G C G T C G T A T G C G A G T C A T A G A
          T C C C T T C A C G A C T G A T C G A C T A
          C G A _ _ _ _ _ 3'
```

Using the SEQAIDII programme, restriction enzyme analysis, search for TATA box, and ORF and comparison of pRG4 221 bp partial sequence with 5'sequence of rice glutelin genomic clone (10) was carried out.

Restriction endonucleases site analysis shows the presence of single site for some enzymes like AccI,

AluI, AvaI, EcoRII, FokI, HaeIII, HpaII, MspI, SmaI, XhoII, XmaI, two sites for HhaI, HinfI, NciI three sites for DpnI, MboI, MnlI, Sau3AI and four sites for TaqI. Further analysis shows that this sequence has G+C content of about 52.9%. There are three open reading frames at positions 114 bp, 127 bp and 183 bp and TATA box is at -27 bases upstream to the ORF starting at 127 bases. In most of the seed storage protein genes from monocots, TATA boxes are at about -30 bases upstream from the transcription start sites. (8,9). In case of rice glutelin TATA box is found -28 bp upstream of transcription start site (7).

Finally 221 bp pRG4 sequence was compared with 1st 500 bases of rice glutelin genomic sequence of Takaiwa et al (10). The sequence shows considerable homology with the 5' flanking region of the glutelin gene.

REFERENCES

1. Shure,M., Wessler,S., Federoff,N. (1983)
   Cell,35, 225.
2. Rosenberg,S.M., Stahl,M.M., Kobayashi,I., Stahl,F.W. (1985)
   Gene, 38, 165.
3. Benton,W.D. and Davis,R.W.(1977)
   Science,196, 180.
4. Takaiwa,F., Kikuchi,S., Oono,K. (1987)
   Mol.Gen.Genet., 208, 15.
5. Co-Shine Wang, Shastri,K., Wen,L., Huang,J.K., Sonthayanon,B., Muthukrishnan,S.,Reeck,.R.(1987)
   FEBS Lett., 222, 135.
6. Short,J.M., Fernandez,J.M., Sorge,J.A., Huse,W.D. (1988)
   Nucl. Acids Res., 16, 7583.
. Okita,T.W., Hwang,Y.S., Hnilo,J. Kim,W.T., Aryan,A.P, Larson,R. and Krishnan,H.B. (1989)
   J.Biol.Chem., 264, 12573.
8. Forde,J., Malpica,J-M., Halford,N., Shewry,P.R., Miflin,B.J., Anderson,O. and Geene,F.C. (1985)
   Nucl. Acids Res., 13, 7327.
9. De Rose,R.T., Ma,Din-Pow, Kwen,I.S., Hasnain,S.E., Klassy,R.C., Hall,T.C. (1989)
   Plant Mol.Biol., 12, 245.

10. Takaiwa,F., Ebinuma,H., Kikuchi,S. and Oono,K. (1987)
FEBS Lett., 221,43.

CHAPTER IV

DISCUSSION

Discussion

4.1 Genetic manipulations of seed storage protein genes: A multidisciplinary approach

Genetic manipulation techniques have progressed rapidly during last 10 years and now it is possible to produce plants with altered genotypes and phenotypes. The most immediate application of genetic manipulation techniques has been the production of new plant varieties that can ward off diseases, damage by insects or unwanted effects from herbicides. Apart from these specific examples several other plant genes are also potential targets for plant genetic engineering. These include genes for improvement of quality of seed storage proteins, enhanced photosynthetic efficiency, improvement in symbiotic and N_2 fixing capacity and so on. Seed storage proteins are especially the major targets for genetic engineering since they are important in human nutrition. Todate seed storage proteins and their genes have been studied in a number of cereals such as wheat, maize, rye, oat and barley. In rice, however, these studies have been mainly restricted to the japonica subspecies while very little information is available on indica cultivars which are widely consumed in India and neighbouring countries.

In our laboratory, the project on "Genetic manipulation of seed storage proteins in indica rice" has been initiated since last 3 years. This project is considered to be a multidisciplinary one and work in following directions has been undertaken :-

(1) A detailed structural study of glutelin storage proteins.

(2) Isolation and characterization of glutelin genomic clones.

(3) Isolation and characterization of cDNAs encoding glutelin.

(4) Synthesis, transport and deposition of glutelin storage proteins.

(5) Modification of glutelin gene by in vitro mutagenesis.

(6) Development of a suitable transformation system for studying expression of native and modified glutelin gene.

In general it is thought that designing of storage proteins with specific functional attributes without adverse effects on seed biology requires thorough knowledge about the molecular basis of functional properties of seed proteins and the critical features of seed protein structures that are essential in post translational processing and packaging.

I was specifically involved with the isolation and

characterization of glutelin gene from Basmati 370 variety of rice.

4.2 Selection of vectors for construction of genomic library

Although genes with some background information can be isolated by most recent approaches such as PCR technology (1), genomic libraries are still frequently constructed as a routine method for cloning of genes. I have constructed a genomic library of Basmati-370 in a Charon40 vector. This work turned out to be a marathon effort in a developing country like India, since I could not use the ready made vector kits and packaging extracts.

Earlier several plant genomic libraries have been constructed using lambda replacement vectors such as EMBL vectors and Charon vectors for isolation of seed storage proteins genes (TABLE 4.1).

The lambda derived EMBL3 and EMBL4 vectors have been designed to reduce the number of non recombinant background phages that can accept DNA inserts from 9-23 kbp in length (2). The multiple cloning sites are located between the lambda arms and the stuffer fragment. Double digestion of the vectors prevents the stuffer fragment from religating to the arms.

TABLE 4.1

Vectors and probes used for cloning seed storage protein genes in cereals

Plant	Protein	Vector used	Probe used	Reference
Wheat	alpha/β gliadins	Charon35	-	3
	gliadin	Charon32	cDNA pT10A10	4
	HMW Prolamin	Charon35	cDNA	5
	LMW glutenin	Charon35	LMW glutenin cDNA pTag544	6
Sorghum	Kafirin	Charon35	cDNA pSKR2, pSKR8 synthetic oligonucleotide	7
Barley	B hordeins	Charon32	cDNA pB7, pB11	8
Maize	Z-19	Charon4A	cDNA pZ19.1	9
	Z-22	LambdaL47.1	cDNA pCM6, pCM41	10
	Z-21	Charon35	incomplete genomic Z 22-21 clone	11
Rice	glutelin	EMBL 3 Charon27	pREE61 cDNA insert as a probe	12
	glutelin	Charon35	pG22 glutelin cDNA	13
	glutelin	Lambda2001	Oligoprobe glu-1 Oligoprobe glu-2 & PCR fragment pAGL of glutelin gene	14

Recently developed lambda DASH II vector is similar to EMBL3 and EMBL4 vectors, but includes T3 and T7 RNA polymerase promoters, as well as NotI sites flanking the multiple cloning site (2). These promoters can be used to make end specific RNA probes to screen libraries for overlapping genomic clones and can be used as target sites for oligonucleotide hybridization and mapping.

Charon derivatives have been constructed by Blattner F.B. et al (1977) (15-17) by mutations that made them simpler to use and safe for biological containment of recombinant DNA. Since the size of insert accommodated by these vectors is in the range of 5-20 kbp,the labour of screening large number of clones is reduced. In my initial phase of work, I used Charon 35 as vector for construction of rice genomic library. However, I could not make progress in this work due to several problems like degradation of DNA during isolation, poor DNA yields and lack of digestibility of DNA with restriction enzymes. I, therefore, decided to use Charon 40 vector which was the latest construct in Charon series. This vector has some marked advantages over earlier series of Charon vectors. It is well known that the repetitive DNA sequences cloned in earlier charon vectors can undergo recombinations in $recA^{+}$ host and could be lost from the library.

Charon40, on the other hand, can use recA⁻ *E. coli* strains as host which facilitates the stability of repetitive DNAs. This asset of Charon 40 vector was of great significance in getting a complete representation of rice genome in the library since rice has about 52 % of the total genomic DNA as repetitive DNA (18).

Since Charon40 and its arms were not commercially available, I decided to isolate Charon40 DNA from the lysates and purify the arms. I encountered some problems in isolating Charon 40 DNA also. Here I realized that the purity of chloroform and PEG was extremely important and a change in batch of these chemicals created a lot of trouble in getting the phage DNA without any degradation. Inspite of all these difficulties, I could finally succeed in getting Charon 40 DNA which was good in its quality and quantity.

4.3 Preparation of rice DNA and packaging extract for library construction

Isolation of high molecular weight DNA and its digestion with a suitable restriction enzyme to get a proper size range is an extremely important step in construction of a genomic library. Plant cells invariably contain a lot of phenolic compounds and polysaccharides which have to be removed with a

minimum shearing of DNA. I tried different methods for DNA extraction (19-21) and among these, the method of Shure et al (22) worked well in our laboratory.

Here the extraction buffer contains 8 M urea which removes all the proteins and polysaccharides. Using this method, I could isolate rice DNA of size above 50 kbp.

The construction of a completely random genomic DNA library can only be obtained if the target DNA is fractionated by a sequence independent method such as mechanical shearing. However, the cloning of sheared target DNAs is not straightforward because many additional enzymatic steps are required to attach RE sites linkers to facilitate cloning. It is more convenient to fractionate target genomic DNAs by using a series of partial restriction enzyme digestions followed by a physical separation of fragments in the size range acceptable to the cloning vector. Probability of getting the intact gene is more when partially digested genomic DNA is used for the preparation of library.

In recent years, packaging extracts are available at reasonable cost from commercial sources such as Stratagene, USA (Giga pack Gold) which work extremely well with very high efficiencies such as 5 x 10^8 PFU/ug DNA. However, it was difficult to procure these kits

when I began my work in 1988. There was, therefore, a need to standardize procedure for preparation of a good, efficient packaging extract. Of many protocols available for the preparation of packaging extracts (23,24), I used the latest protocol where only single lysogenic strain E. coli SMR10 (25) is used and only one extract is prepared. Extracts made in this way have a lower background of nonrecombinant plaques than the classical binary mixture. Several attempts were made for preparing packaging extract over the period of 6-8 months. Here, I noticed that if SMR 10 culture took more than 2.5 -3 hr to achieve the desired O.D., the efficiency of packaging was less. To get this particular O.D. as fast as possible, initial shaking of the culture in LB medium at 300 rpm at 34°C was found to be extremely essential. Similarly in the protocol it was mentioned that the frozen packaging extract should be thawed at 4°C for 2-3 min before adding DNA in it. When I followed the same instructions, I could not get the desired efficiency . After some trails, I,however, could succeed in getting a good packaging efficiency of about 1×10^8 PFU/ug lambda DNA when thawing of packaging extract and mixing of phage DNA with it were carried out simultaneously.

Using this packaging extract the rice genomic

library was constructed which represented 7-8 genomes of rice.

4.4 Identification of genomic glutelin clone

Genomic libraries which contain the clones representing each fragment of the genome are generally screened using several types of nucleic acid probes to fish out the desired clone. Three types of nucleic acid probes are usually used. Homologous probes like cDNA or partial clones identified previously could be used at very high stringencies. Heterologous probes such as same gene cloned from another species or a related gene cloned from the same species could be used as a probe by performing a series of Southern hybridizations at different stringencies. Oligonucleotide probes are synthetic sequences that are synthesized in vitro. Three types of oligonucleotide probes are in common use: 1) Single oligonucleotide probe of defined sequence 2) Pools of short oligonucleotide probes whose sequences are highly degenerate and 3) Pools of longer oligonucleotides of lesser degeneracy.

In my work,I have used a single oligonucleotide of defined sequence which was synthesized based on the glutelin cDNA sequence from japonica variety of rice (26). A similar type of synthetic oligonucleotide probe was used for isolation of genomic clone encoding

kafirin from the sorghum genomic library (7). Another example is the use of oligonucleotide probes glu-1 and glu-2 to screen rice genomic library in vector Lambda 2000 (14) for isolation of glutelin clones.

4.5 Structural features of glutelin genes in rice

When my work was undertaken, only a few reports regarding glutelin cDNA cloning were available from different varieties of rice.

I have isolated genomic clone (lambda RG4) encoding glutelin from the genomic library of Basmati rice in Charon40 vector. The glutelin genes can be classified into several subfamilies according to their DNA sequences. The Gt1 and Gt2 subfamilies are closely related and Gt3 subfamily is different from the two mentioned above (13). Similar type of glutelin gene belonging to Gt2 family has been characterized (14). Takaiwa et al (12) have shown that rice glutelin cDNAs could be classified into two groups on the basis of physical maps and sequences. Two dimensional gel electrophoretic analysis of rice glutelin has revealed that acidic and basic subunits can be resolved into at least 12 and 9 polypeptides respectively (27) which possibly are encoded by different glutelin gene subfamilies. Each of the subfamilies contains about 5-8

copies per haploid rice genome. I have observed 5-8 copies of glutelin gene in basmati genome. In Southern hybridization using pRG4, two or more hybridizing bands are seen which probably represent different glutelin subfamilies in basmati genome. Although there are several reports of nucleotide sequencing of glutelin cDNA clones (13,28-30) , there are very few reports on sequencing of glutelin genomic clones (12-14). In all comparisons, however, the sequence identity of the exons of all the three genes is much more conserved (4-6% divergence) than that of flanking and intron regions. Presumably, the primary sequences of the glutelins have been maintained for correct folding and post translational processing of the polypeptide and aggregation into descrete organelles in seed tissue.

The 5' flanking region exhibits a considerable homology among all Gt genes. In addition, three short conserved stretches of sequences upstream of the translational start sites are shared by all glutelin genes. A conserved TATA box is located 28 bp from the major transcriptional initiation site. Consistent with the strong homology in exons, the promoter sequences are also highly conserved. The presence of short inverted repeats is reminiscent of transposable elements, suggesting the involvement of rice glutelin genes in evolution. Multiple polyadenylation signals

are seen in glutelin genes. Three small introns which are relatively A-T rich ranging in size from 75-111 bp are observed. All of the intron borders obey the AG/GT rule (13). To understand the basic mechanism of tissue specific gene expression, transient expression assays using immature rice seeds have been done which show that its 5' flanking sequence can direct the synthesis of β glucuronidase (GUS) when fused to the upstream sequence of the GUS coding region (32).

4.6 Future perspectives

In summary, knowledge of structure, organization and regulation of gene brings us closer to actual applications of genetic engineering to improve the nutritional properties of grains. Today molecular biologists have the dream of designing seed proteins with specific functional attributes without adverse effects on seed biology.

Cereals are the major sources of protein worldwide and hence it will be very beneficial if the protein content and quality of cereals is improved. The protein content of cereals varies from 7% - 14%. Cereals are deficient in essential amimo acids like lysine, tryptophan and threonine (32). The high content of prolamins is responsible for this low content of

essential amino acids in cereals. Rice and oats have a better balance of essential amino acids than other cereals due to a lower content of prolamins (33). Improving the protein quality of rice grain will bring additional significant benefits because the N digestibility and N retention of rice diets are low (34).

The number and complexity of seed storage protein genes present a problem if mutations are to be made to modify the amino acid composition of the proteins they code for. The most strongly expressed gene would be modified to bring about a significant change in the overall amino acid composition of seed. It is possible to enhance seed specific expression significantly by inserting a seed specific DNA sequence into a strong constitutive promoter such as cauliflower mosaic virus (CaMV) 35 S promoter. Comparison of gene sequences has revealed sequences that may be useful in genetic engineering work . These include putative seed specific promoters, quantitative control sequences and variable protein sequences where gene modifications can conceivably be made without affecting synthesis, transport and deposition of these proteins in the seed.

The functionality of proteins is very important and needs to be maintained. Computer modelling of structure-function relationships in proteins, mainly

enzymes, is increasingly being used to design proteins rationally for specific funtions. Such an approach could be used to determine how site-specific mutations would affect functional properties of seed proteins. Such information, combined with a knowledge of the critical sequence of seed proteins needed for post translational processing and packaging, would allow the rational design of proteins with specific functional attributes without adverse effects on seed biology.

REFERENCES

1. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988)
   Science, 239, 487.
2. Frischauf, A., Lehrach, H., Poulstka, K. and Murray, N. (1983)
   J. Mol. Biol. 170, 827.
3. Sumner-Smith, M., Rafalski, J.A., Sugiyama, T., Stoll, M. and Soll, D. (1985)
   Nucl. Acids Res. 13, 3905.
4. Anderson, O.D., Litts, J.C., Gautier, M-F. and Greene, F.C. (1984)
   Nucl. Acids Res. 12, 8129.
5. Forde, J., Malpica, J.M., Halford, N.G., Shewry, P.R., Anderson, O.D., Greene, F.C. and Miflin, B.J. (1985)
   Nucl. Acids Res. 13, 6817.
6. Colot, V., Bartels, D., Thompson, R., and Flavell, R. (1989)
   Mol. Gen. Genet. 216, 81.
7. DeRose, R.T., Ma, D-P., Kwon, I-S., Hasnain, S.E., Klassey, R.C., Hall, T.C. (1989)
   Plant Mol. Biol. 12, 245.
8. Forde, B.G., Kreis, M., Williamson, M.S., Fry, R.P., Pywell, J., Shewry, P.R., Bunce, N. and Miflin, B.J.

(1985)

EMBO J. 4, 9.

9. Pederson, K., Devereux, J., Wilson, D.R. and Larkins, B.A. (1982)

Cell, 29, 1015.

10. Spena, A., Viotti, A. and Pirrotta, V. (1983)

J. Mol. Biol. 169,799.

11. Kridl, J.C., Vieira, J., Rubenstein, I. and Messing, J. (1984)

Gene, 28, 112.

12. Takaiwa, F., Ebinuma, H., Kikuchi, S. and Oono, K. (1987)

FEBS. Lett. 221, 43.

13. Okita, T.W., Hwang, Y.S., Hnilo, J., Kim, W.T., Aryan, A.P., Larson, R. and Krishnan, H.B. (1989)

J. Biol. Chem. 264, 12573.

14. Kim, S.Y. and Wu, R. (1990)

Nucl. Acids Res. 18, 6845.

15. 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., Sheldon, E. L. and Smithies, O. (1977)

Science, 196 ,161.

16. Dunn, I.,S. and Blattner, F. R. (1987)

Nucl. Acids Res. 15, 2677.

17. Loenen, W. A. M. and Blattner, F. R. (1983)

Gene, 26, 171.

18. Deshpande, V. G. and Ranjekar, P. K. (1980)
Hoppe - Seylers Z. Physiol. Chem. 361, 1223.

19. Ranjekar, P. K., Pallotta, D. and Lafontaine, J. G. (1976)
Biochim. Biophys. Acta. 425, 30.

20. Marmur, J.(1961)
J. Mol. Biol., 3, 208.

21. Murray, M. G. and Thompson, W. F. (1980)
Nucl. Acids Res. 8, 4321.

22. Shure, M., Wessler, S. and Federoff, N. (1983)
Cell, 35, 225.

23. Sternberg, N., Tiemeier, D. and Enquist, L. (1977)
Gene, 1, 255.

24. Hohn, B. (1979)
Methods Enzymol. 68, 299.

25. Rosenberg, S.M. (1988)
Gene, 39, 313.

26. Takaiwa, F., Kikuchi, S. and Oono, K. (1986)
Mol. Gen. Genet. 208, 15.

27. Wang, C.-S., Shastri, K., Wen, L., Huang, J.-K., Sonthayanon, B., Muthukrishnan, S. and Reeck, G.R. (1987)
FEBS Lett. 222, 135.

28. Takehiro, M., Kunitomo, K., Yoshimori, S., Norihiro,

M., Hibino, T., Tanaka, K. and Fujii, S. (1989).
Plant Mol. Biol. 12, 723.

29. Higuchi, W. and Fukazawa, C. (1987)
Gene, 55, 243.

30. Takaiwa, F., Kikuchi, S. and Oono, K. (1989)
Nucl. Acids Res. 17, 3289.

31. Leisy, D.J., Hnilo, J., Zhao, Y. and Okita, T.W. (1990)
Plant Mol. Biol. 14, 41.

32. Eggum, B.O. and Beames, R.M. (1983)
In Seed Proteins: Biochemistry, Genetics, Nutritive value., Eds. Nijhoff. p499.

33. Shewry, P.R. and Miflin, B.J. (1984)
Adv. Cereal Sci. Technol., 7, 1.

34. Roxas, B.V., Intengan, C.L., and Juliano, B.O. (1979)
J. Nutr., 109, 832.