The molecular biology of self-incompatibility in Brassica

Scutt, Charles Patrick

The Molecular Biology of Self-Incompatibility in *Brassica*

A thesis submitted by Charles Patrick Scutt B.Sc. (Reading) in accordance with the requirements for the degree of Doctor of Philosophy in the University of Durham.

Department of Biological Sciences. January 1990

MEMORANDUM

Part of the work presented in this thesis has been accepted for publication as follows:

Scutt CP, Gates PJ, Gatehouse JA, Boulter D, Croy RRD. A cDNA encoding an S-locus-specific glycoprotein from *Brassica oleracea* plants containing the S_5 self-incompatibility allele. Molecular and General Genetics (in press).

STATEMENT

No part of this thesis has previously been submitted for a degree in this or any other university. I declare that unless otherwise indicated, the work presented herein is entirely my own.

ACKNOWLEDGEMENTS

I am most grateful to my supervisors: Dr. R. Croy, Dr. P. Gates and Professor D. Boulter for their excellent supervision.

I am indebted to the following people for much helpful advice and many interesting discussions: Dr. John Gatehouse, Dr. Marta Evans, Dr. Vaughan Hilder; Dr. Glyn Edwards and Dr. Tony Bradshaw (Shell Research, Sittingbourne, Kent), Mr. Andy Thompson and Mr. Tony Fordham-Skelton.

I would like to thank the following people for technical assistance: Dr. John Gatehouse (nucleotide sequence analysis), Miss Julia Bryden and Mr. Paul Preston (nucleotide sequencing reactions) and Mr. John Gilroy (oligonucleotide synthesis).

Thanks are due to the following for the supply of biological materials: Dr. David Ockendon (IHR, Wellsbourne), Dr. Chris Franklin (Birmingham University), Mr. Simon Linington (RBG Kew Seed Bank, Wakehurst Place, Sussex), Mr. Andy Thompson and Mr. Sami Yaisch.

I am grateful to Ms. Shirley Stewart for typing the manuscript and to Mr. David Hutchinson and Mr. Paul Sidney for photographic work.

I acknowledge receipt of an SERC biotechnology research studentship.

ABSTRACT

The Molecular Biology of Self-Incompatibility in *Brassica*

Charles P. Scutt.

Self-incompatibility in *Brassica* is controlled by a single "S"-locus. Different self-incompatible genotypes of *Brassica* are correlated with the occurrence of S-locus-specific glycoproteins (SLSGs), cDNA sequences for certain of which have been determined.

In this thesis two cDNAs are presented, derived from a single homozygous line of *Brassica oleracea*, which are both partially homologous to known SLSG genes. One of these S-like sequences is possibly a novel form of the SLSG gene (corresponding to the S_5 allele), whilst the other appears invarient between *Brassica* lines containing different S-alleles (by comparison with published sequence date). The organisation of genomic sequences homologous to both classes of S-like cDNAs was investigated by Southern blotting using genomic DNA extracts from *Brassica* lines containing a range of S-alleles. The expression of genes homologous to the two cDNAs has been investigated by Northern hybridization and found to be specific to stigma tissue; no transcripts of S-like sequences are detectable in anther tissue over the developmental period during which expression of male S-gene sequences might be expected.

The existence of S-like sequences, homologous to those from *Brassica oleracea*, has been investigated by Southern hybridization using genomic DNA extracts from a range of *Brassica* species and other related species. Inferences are drawn from this as to the timing of the gene duplication event which led to multiple S-like sequences, in relation to the timing of the evolutionary divergence of related taxa. The divergence of S-like sequences between different lines of *B. oleracea* has been analysed by comparison of DNA sequence data presented in this thesis with published sequence data. The relative rates of evolution of two domains within S-like gene sequences was estimated.

A further three unrelated classes of tissue - specifically expressed sequences have been selected from a *Brassica* stigma cDNA library. One of these encodes a glycine-rich cell wall protein; another represents an unidentified gene showing stigma specific expression; the third represents an unidentified gene showing expression in stigma, style and anther tissue, though not in leaf tissue. cDNA sequence data, genomic Southern hybridizations and Northern hybridizations are presented for all three of these classes of cDNAs.

ABBREVIATIONS

ANS	8-anilino-1-naphalinesulphonic acid
ATP	adenosine triphosphate
b	bases
bp	base pairs
cDNA	complementary DNA
cv	cultivar
dCTP	deoxycytidine-5'-triphosphate
DNA	deoxyribonucleic acid
GRP	glycine-rich protein
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
IPTG	isopropyl-2-D-thiogalactoside
Kb	kilobase pairs
KDa	kiloDaltons
MES	2(N-morpholino)ethanesulphonic acid
mRNA	messenger RNA
MW	relative molecular weight
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
POPOP	1,4-bis(5-phenyl-2-oxazolyl)benzene
PPO	2,5-diphenyloxazole
RNA	ribonucleic acid
RNase A	ribonuclease A
rRNA	ribosomal RNA
SDS	sodium dodecyl sulphate
SI	self-incompatibility
SLSG	S-locus-specific glycoprotein
SSC	standard saline citrate
tRNA	transfer RNA
UV	ultraviolet
var.	variety
X-Gal	5-bromo-4-chloro-3-indolyl-2-D-galactoside

LIST OF FIGURES

CONTENTS

CHAPTER ONE

INTRODUCTION

INTRODUCTION

1.1 General Introduction

1.1.1 The Definition and Natural Function of Self-Incompatability

Self-incompatability (SI) is a powerful adaptation to outbreeding found in many species of angiosperms. Lundqvist (1964) defines SI as:-

"the inability of a fertile hermaphrodite seed-plant to produce zygotes after self-pollination".

SI comprises a molecular recognition system in which genotypically specified recognition molecules borne in, or on pollen grains and also within the style, or on the stigmatic surface, interact to arrest the post-pollination development of self-pollen. It represents a form of intercellular recognition between plant cells and may act as a paradigm for other such systems operating in: interspecific-incompatibility, pathogenesis, symbiosis and gametic fusion.

The natural function of SI systems seems clear. Inbreeding leads to homozygosity and genetic fixation; recessive alleles are lost from the gene pool and recombination has little or no effect in re-sorting the allelic forms of linked genes. In well adapted plant species this need not matter. Indeed, the great saving of energy in predominantly inbreeding species in not attracting pollination vectors with showy or scented flowers and not producing a reward in the form of nectar or excess pollen, is considerable. This may be highly advantageous and must

surely be one important reason for the frequent natural selection of revertants to the self-compatible state. However, at times of rapid environmental change the continued existence of a species depends on its capacity to adapt, which relies in turn on a source of allelic diversity and the ability, through outbreeding, to reshuffle genes such that the most favourable characteristics are expressed together and preserved.

There are several types of outbreeding adaptation which have been adopted by various angiosperm groups, but amongst these SI is both the most subtle and powerful. SI does not suffer from the inherent reduction in fecundity associated with dioecism. Also, unlike dioecism, SI can sometimes allow self-fertilization to occur as an "emergency measure". Such a measure may be required if, for example, appropriate pollination vectors are temporarily scarce. SI systems may allow some degree of selfing by: certain flowers, especially those produced late in the flowering season, being self-compatible; reversion to the self-compatible state in a proportion of the population; and particular combinations of weak alleles, which control SI, allowing self-fertilization. The outbreeding adaptations of protandry, protogyny, and monoecism are incomplete barriers to self-fertilization, but are obligately so, lacking the plasticity of SI in this regard.

1.1.2 The Diversity and Genetics of Self-Incompatibility

SI may be conveniently divided into a number of major categories. Inevitably, such categorization implies a degree of homology between systems included within any

group, though such an assumption may be erroneous. The categorisation used here is standard and is based upon the nature of the genetical control of the SI reaction of pollen and the existence, or otherwise, of distinct floral morphologies which correlate with the SI genotype.

The most frequent system of SI is homomorphic gametophytic monofactorial SI (or just 'gametophytic SI'). It is called 'gametophytic' because the genotype of the pollen grain (which is the gametophytic generation) determines its own SI recognition reaction. Early investigations into the genetics of gametophytic SI systems were performed by East and Mangelsdorf (1925) and others (reviewed by de Nettancourt, 1977). Such systems were found to operate with a single locus, (termed the S-locus), which exists in a polyallelic series ranging into hundreds of allelic forms in some species. Identity of the S-allele present in a pollen grain with one of those present in the style leads to the abortion of pollen tube growth some distance into the stylar transmitting tissue, though some gametophytic systems have an almost immediate effect . S-alleles operating in the style show strict co-dominance in gametophytic systems. In a plant which has undergone a doubling of its chromosomes, the SI system of the (tetraploid) stigma remains operative whilst that of the (diploid) pollen breaks down. Gametophytic SI systems are usually correlated with the occurrence of: a "wet" stigma system, binucleate pollen, distinctly organized stylar transmitting tissue and biphasic growth of compatible pollen tubes; (the tubes grow more slowly at first, perhaps implying a stage of selection against self-pollen).

Other systems, most notably in the Gramineae, exhibit gametophytic control of pollen SI, but this is accomplished through identity of alleles at two separate loci (S and Z) in both stigma and pollen. The inhibition of tube growth in self-incompatible grasses occurs at the surface of the dry, feathery stigmas.

Homomorphic sporophytic SI (or sporophytic SI for brevity) is found in only six angiosperm families to date, (Charlesworth, 1988 and references therein). It can be contrasted with gametophytic SI in a number of respects, although both systems are determined by a single polyallelic S-locus. In this type of SI system the diploid genotype of the parent plant specifies the SI phenotype of the pollen it produces, such that a single grain may contain two separate self-incompatible identities. This is complicated by the existence of dominance relationships in both pollen and stigma, most thoroughly investigated in *Brassica* (Ockendon, 1975), where over fifty alleles are known. These dominance relationships range from codominance to complete dominance and may be different in pollen and stigma.

The current theory concerning the origin of the pollen-borne components of SI in sporophytic systems is that they are produced by the tapetum (Heslop-Harrison, 1967) and are applied together with the lipidic pollen coating which invests the outer wall layers. Dickinson and Lewis (1973) demonstrated *Raphanus* pollen wall materials to be capable of eliciting the response of callose formation in stigmatic papillae; a response which occurs naturally on contact with incompatible pollen. The source of such outer-wall-held

materials has been shown to be the tapetum (Dickinson and Lewis, 1973).

Sporophytic systems of SI are correlated with the occurrence of "dry" stigma systems, monophasic growth of compatible tubes and the lack of a differentiated stylar transmitting tissue. Inhibition of tube growth in incompatible pollinations is rapid, occurring at the stigmatic surface. Sporophytic SI systems tend to operate in plants which have trinucleate pollen grains.

Some SI systems work not with a polyallelic series of S-alleles, but with only two, one of which is completely dominant. In these plants a population is divided into homozygous recessives (ss) and heterozygotes (Ss). Expression of the S-gene is sporophytic such that all pollen from a plant which is heterozygous at the S-locus expresses the phenotype associated with the dominant allele. One individual in such a population of plants is both self-incompatible and cross-incompatible with half of the remainder. This system may be regarded as less efficient than the polyallelic forms previously discussed as it provides an extensive random barrier to cross-compatibility.

The S-genotype in these diallelic systems is almost invariably correlated with some form of floral heteromorphism and hence these systems are referred to as heteromorphic SI systems. In the distylic species *Primula*, for example, lengths of styles and filaments correlate with the constitution of the S-locus (Darwin, 1877). This structural dimorphism is adapted for the deposition of pollen on the proboscis of a bee in the correct position for transfer to the stigma of a compatible, but not an

incompatible flower. Such heteromorphism is regarded as a physical reinforcement of the SI system and the genes which control the relative lengths of the style and filaments are closely linked at the S-locus.

Certain heteromorphic species, for example in the Lythraceae and Oxalidaceae, exhibit a floral trimorphism. In these species the SI phenotype and floral structure is governed by genes clustered in two allelic forms at two separate loci.

A number of other SI systems in existence display characteristics which differ from those discussed previously. Certain systems operate with more than two loci controlling SI. In others pollen tube development is arrested at fertilization, or SI is effected by the abortion of zygotes resulting from self-fertilization. These systems are discussed by de Nettancourt (1977).

1.2 The Evolution of Self-Incompatibility

SI is known in over half the families of flowering plants (East, 1940) and in the vast majority of cases investigated, a gametophytic multiallelic system is operating. Whitehouse (1950) proposed the gametophytic SI system to be ancestral and to be present in the single progenitor species from which the angiosperms are purported to have arisen. Furthermore, Whitehouse contends that all present-day gametophytic SI systems arose by modification of the ancestral system and attributes the success of the angiosperms, in part, to the possession of SI. Most of the gametophytic SI systems examined do appear to operate in a similar manner. They correlate with a number of anatomical,

cytological and physiological factors which appear absent in other SI systems.

Brewbaker (1957), Crowe (1964) and others support the view of a single progenitor gametophytic system and suggest that this must have been multiallelic; a diallelic system is regarded by these authors as being of insufficient selective advantage over other outbreeding mechanisms in accounting for the rapid spread of the angiosperms in becoming the dominant land plants.

Accepting, for the moment, the occurrence of a single progenitor gametophytic SI system, the question remains as to the origin of sporophytic, heteromorphic and other SI systems. De Nettancourt (1977) reviews the literature concerning speculative schemes for the evolution of the various known SI systems from a gametophytic system, though no evidence exists to support such descent. The development of relatively unsophisticated systems such as diallelic SI from a powerful multiallelic SI system would appear, in any case, counterintuitive.

Sporophytic homomorphic SI systems have been found in only six families to date (reviewed by Charlesworth, 1988), and these are widely dispersed through conventional taxonomic schemes. It would appear, therefore, that sporophytic SI systems may have arisen a number of times in the progenitors of the extant groups which currently exhibit them. It is not at all clear that sporophytic SI systems arose by adaptation of gametophytic systems; they may have arisen in groups which had lost SI, by modification of physiological mechanisms unconnected with incompatiblity phenomena.

Gibbs (1986) points out the lack of data upon which many elaborate evolutionary schemes, accounting for present-day systems of SI, are based. After lengthy criticism of various hypotheses which argue, seemingly coherently, towards radically different conclusions, Gibbs (1986) can find no evidence to link homomorphic sporophytic systems with heteromorphic ones (though the latter also exhibit sporophytic control). Gibbs (1986) can also find no compelling reason for disputing the claim (Whitehouse, 1950) that a gametophytic system was present in the progenitor species of the angiosperms, but equally, no evidence exists to support a link between this and any sporophytic or heteromorphic system. This view is supported by the molecular biological data which has emerged recently. Though sequence data from only the Cruciferae and Solanaceae is currently available, this tends to suggest separate evolutionary origins for sporophytic and gametophytic systems.

The occurrence of gametophytic and sporophytic SI can be correlated with that of wet and dry stigma systems, respectively. A mechanism has been suggested by Professor H.G. Dickinson (personal communication) as to how the evolution of the sporophytic SI system, (either from a gametophytic system or *de novo*), may have occurred in parallel to the evolution of the dry stigma. This hypothesis argues that wet stigmas are disadvantageous as they provide a moist and nutritive environment for colonisation by plant pathogens. The dry stigma is proposed to have evolved as an adaptation to the avoidance of pathogenesis. Coincidentally, the secretory cells at the

stigmatic surface are proposed to have synthesised a range of glycoproteins. In anthers, the analogous secretory tissue to the stigmatic surface is the tapetum and it is hypothesised that these two tissues might interpret gene promoters, (or higher level genetic control elements), in a similar manner. In this way stigmatic glycoproteins might become synthesized in and secreted from the tapetum. The adaptation of such glycoproteins to act in cellular recognition would provide one essential element of a sporophytic SI system.

Sporophytic SI systems, evolved in the manner outlined above, may have been superimposed upon a previously functioning gametophytic SI system which was no longer completely congruous with the changed stylar anatomy after the evolutionary development of the dry stigma. Certain evidence has been amassed (Lewis *et al.*, 1988) and Zuberi and Lewis, 1988) that in *Raphanus* and *Brassica*, a gametophytic SI system is operating in addition to the well known sporophytic SI system. It would be interesting to compare the active recognition component of the hypothesized gametophytic system of these plants (termed the G-gene), with the putative S-genes cloned from gametophytic systems such as *Nicotiana*. No homology has been found between sporophytic and gametophytic S-genes, but homology may exist between G-genes from sporophytic systems and gametophytic S-genes.

Some light will, undoubtedly, be shed on the evolution of SI by molecular biological data. Sequence information for S-genes provides direct answers to the question of homology between SI systems. (This, of course, depends upon

the correct identification of the S-gene sequences in the groups to be compared.) Furthermore, "Southern" blotting may be used to test whether sequences related to an S-gene from one species are present in another. This technique provides less reliable data than direct sequence comparison as it is not possible to directly surmise the function of a genomic sequence which hybridizes to an S-gene probe.

1.3 The Physiology of Self-Incompatibility

The SI reaction may be conceptually reduced to a recognition event between a self-pollen grain and the female reproductive tissues, followed by a mechanism leading to the rejection of the self-pollen.

The molecules involved in recognition have been the subject of much recent research and detailed discussion of these is deferred to the next section. Initially, the whole process of signal transduction and response will be considered.

In gametophytic SI systems the stigma is usually wet and there is no barrier to self-pollen hydration and germination. Self-pollen tube elongation is arrested within the transmitting tissue of the style, the length to which incompatible tubes grow being a function of the strength of the S-alleles present (Mulcahy and Mulcahy, 1988). Recognition of "self" in gametophytic SI systems acts only to arrest pollen tube growth, in contrast to a multiplicity of responses which occur in sporophytic systems. Experiments with *Nicotiana* (Hoggart, 1984) demonstrate the putative stylar S-gene product to inhibit the tube growth of *in vitro* germinated pollen. This may be interpreted as

suggesting the only requirement of the style in effecting SI in *Nicotiana* is the expression of the S-gene in the production of the recognition molecule. All of the ensuing physiology, leading to the cessation of pollen tube growth, occurs within the developing pollen tube.

The inhibition of pollen tube growth might be brought about by a number of factors which have been shown to operate in pollen grains, though evidence of the actual sequence of events in any system remains to be established. Polya *et al.* (1986) have shown germinated *Nicotiana* pollen to contain both Ca^{2+}-dependent and Ca^{2+}-independent protein kinase activities capable of regulating phosphorylation of a number of endogenous polypeptides. Calcium ion concentration is a critical factor in the growth of *Nicotiana* pollen *in vitro* and its regulation may, therefore, be involved in the SI response mechanism.

Cyclic AMP is known in plants (Amrhein, 1977) and has been found to induce pollen tube elongation in *Tradescantia* (Malik *et al.*, 1976) when exogenously applied. Katsumata *et al.* (1978) have noted adenylate cyclase activity in the pollen of *Pinus* and recently some work has focussed on the regulation of cyclic AMP in germinating pollen as a paradigm for signal transduction in gametophytic incompatibility responses (Rougier *et al.*, 1988), the system investigated being that of interspecific gametophytic incompatibility in *Populus*. Clarke *et al.* (1985) have discussed various models of signal transduction in the SI system of *Nicotiana*. They interpret the SI response as the cessation of pollen wall formation and examine a number of potential control points in the biosynthesis of wall polysacharides which could

mediate the effects of self-recognition by stopping this process, though none of these could be specifically implicated.

Sporophytic SI systems exhibit a number of basic differences from gametophytic systems. The mechanism of incompatible pollen tube inhibition may be entirely different in these two types of system. Most of the physiological studies of SI in *Brassica*, an example of a sporophytic system, have been based upon: various levels of microscopic observation; the kinetics of pollen grain hydration on the stigma; and the use of metabolic inhibitors and tracers.

The recognition reaction in *Brassica* determines a number of subsequent events (in contrast to that of gametophytic SI systems). Firstly, Stead *et al.* (1979) have demonstrated that compatible *Brassica* pollen rapidly becomes strongly attached to the stigmatic papillae after pollination, whereas incompatible grains remain easily removable. Following this initial consequence of self-recognition, it has been shown that compatible and incompatible grains may achieve a different hydration status when placed on a stigma (Sarkar *et al.*, 1988), though it has been demonstrated that both compatible and incompatible grains will hydrate to some extent (Elleman and Dickinson, 1986).

A large amount of work has been performed on the ultrastructure of the *Brassica* pollen coating, the stigmatic surface layers, and the events which occur at the interface between these upon pollination. In both compatible and incompatible pollinations Elleman and Dickinson (1986)

observed the pollen coating to fuse with the stigmatic pellicle upon contact and noted a dramatic "coat conversion" in that previously electron-lucent material became highly electron-opaque. This conversion was found to be dependent on the process of hydration on the stigma and does not occur in pollen which hydrates in a moist atmosphere. It also occurs first at the area of the pollen coat which is in contact with the stigmatic pellicle. Some differences in rapidity of "coat conversion" were noted between compatible and incompatible grains. These differences, to some extent, support models of the control of SI in *Brassica* based on the regulation of pollen hydration. Water is the sole requirement of *Brassica* pollen for the activation of germination (Ferrari *et al.*, 1981) and recent work has shown that the supply of water to the pollen grain may be under the control of the stigma (Sarkar *et al.*, 1988). It was found that cycloheximide treatment of the stigma removed its control over pollen hydration so that self-incompatible and interspecifically incompatible grains could rapidly hydrate, as could the compatible controls. Roberts *et al.* (1984) studying the same phenomenon, again using cycloheximide, found the incompatibility system to operate for two to four hours after application of the inhibitor, implying turnover of some polypeptide within that time period which is required for the SI system to function.

Sarkar *et al.* (1988) found tunicamycin, an inhibitor of protein glycosylation, to break down SI in *Brassica* stigmas. This was interpreted as causing its effect through the prevention of glycosylation of the S-gene product, which separate molecular evidence has indicated to be a

glycoprotein. The effect of tunicamycin might, further, be regarded as evidence of the essential role of the oligosaccharide moieties of the S-glycoprotein. Interestingly, the proportion of the S-glycoprotein pool which may be affected by inhibitor treatment, within the time period required to break down SI, is small. This has led Sarkar *et al.* (1988) to suggest that only a proportion of the total pool of S-glycoprotein, possibly that which is most recently synthesised, is active.

In addition to evidence that the stigma may exert control over pollen hydration in *Brassica* and thereby regulate pollen germination as a part of the SI system, there is some evidence that germination inhibitors may be released from stigmas into incompatible pollen grains. Hodgekin and Lyon (1984) have found certain low molecular weight germination inhibitors to be present in *Brassica* stigmas only following self-pollinations, though other classes of such inhibitors were present also in unpollinated and cross-pollinated stigmas. The bioassay used to investigate the effects of these substances was based on the *in vitro* germination of *Petunia* pollen as no such system was available using *Brassica* pollen.

1.4 The Molecular Biology of Self-Incompatibility

The molecular biology of recognition between the female parts of a flower and self-pollen grains has been the subject of much theoretical speculation. Within the last five years molecular biological data has been emerging, though only so far relating to the female component of the

recognition process; for this reason the process of recognition in SI is still not understood.

Early models of recognition were based upon information gleaned from the genetics of SI systems and experiments involving the immunological detection of S-antigens. A model based on the positive stimulation of pollen by the action of dissimilar S-alleles present in pollen and style failed to satisfy the requirements of the system and attention focussed on models based on the oppositional inhibition of like gene products (reviewed by de Nettancourt, 1977).

Lewis (1964) proposed a model of SI recognition based on the formation of multimers between S-gene products. It was proposed that the S-gene is expressed in both male and female tissues such that the single S-determinant is present as a pollen and stylar or stigmatic component. These pools of the S-determinant are supposed to not interact within their native tissue, but at some stage after pollination the products of identical S-alleles associate to create active multimers which inhibit self-pollen tube growth. Lewis (1964) proposed the active inhibitor to be a tetramer formed from homodimers of S-gene products. Lewis' model, despite various apparent theoretical problems, is still a viable concept. Some results will be presented in this thesis which suggest this concept should be modified to allow at least two S-gene sequences, such that male and female-expressed elements are proposed to exist as distinct entities, though closely linked on the chromosome. This refined 'dimer' hypothesis would regard the male and female-expressed determinants of SI to exist in pairs; each

pair would correspond to a single S-allele number. The two sequences proposed would be inherited as a single genetic unit and their gene products would interact after pollination to form active dimers, or multimers, which prevent pollen tube growth. Van der Donk (1975) suggests the male and female-expressed S-gene sequences to be separate in *Petunia hybrida*, supporting this view.

In order to attempt to account for results inconsistent with Lewis' 'dimer' hypothesis, Mulcahy and Mulcahy (1983) proposed the 'heterosis model'. This model, concerned exclusively with gametophytic SI systems, regards the incompatibility reaction as the product of a number of polyallelic loci which each have a regulatory effect on pollen tube growth rate. This hypothesis was heavily criticised by Lawrence *et al.* (1985) and new data obtained by Mulcahy and Mulcahy (1988) has been found to be inconsistent with their earlier model.(The model would predict growth rates of compatible pollen tubes to be slower if strong SI alleles were present, even in compatible pollinations; for *Lycopersicum peruvianum* this was found not to be the case.)

Molecular studies on SI have been performed chiefly on *Nicotiana*, (a gametophytic example), and *Brassica*, (a sporophytic example). The stage of understanding reached in these two model systems is similar at the present time. From both systems a number of allelic forms of a gene have been cloned and sequenced. These genes encode glycoproteins which are heavily implicated as the female-expressed recognition molecules in SI. The chief barrier to further understanding is the lack of information concerning the male

counterpart to the cloned molecules in either of the two experimental systems. It must also be mentioned that, at the time of writing, definitive proof of the role of the cloned sequences as S-genes remains to be established. The evidence for this is marginally stronger in *Nicotiana* where an *in vitro* pollen germination bioassay has been established. The definitive proof of identity required for these putative S-gene sequences must come from transformation studies. Such results have not yet been published.

The molecular studies on *Nicotiana* SI have been pioneered by Professor Clarke and her collaborators in Melbourne. Glycoproteins were identified from stylar extracts of *Nicotiana* whose electrophoretic heterogeneity correlate with particular S-allelic specificities (Anderson *et al.*, 1986). The S-associated glycoprotein corresponding to the S_2 genotype was purified, deglycosylated and N-terminally sequenced. Identification of cDNA molecules encoding this glycoprotein (from a stylar cDNA library) was based upon differential screening with mature and immature stylar cDNA followed by hybridization to a short cDNA probe selectively primed from stylar mRNA by a set of oligonucleotides constructed from the N-terminal sequence of the protein. The cDNA corresponding to the stylar S-associated glycoprotein for the S_2 genotype was fully sequenced.

Subsequent work by the same group has led to the isolation of cDNA clones putatively corresponding to the S_3 and S_6 *N.alata* S-alleles. The homology between the cloned sequences has been shown to be between 63% and 70% at

the amino acid level. The sequences are approximately 200 amino acid residues in length and include a 22 amino acid putative signal peptide. The S_2 sequence shows three potential N-glycosylation sites, but these remain conserved in the other sequences examined, not reinforcing the hypothesis that they are responsible for conferring specificity in the SI reaction.

The *Nicotiana* S-associated glycoprotein gene appears to be present as a single copy sequence with a 94 bp intron. The expression of this gene has been demonstrated by *in situ* hybridization to be limited to the female secretory tissues, namely: the stigma, stylar transmitting tissue and secretory parts of the ovary and epidermal cells of the placenta (Cornish *et al.*, 1987). Maximum levels of message for this gene are detected in the transmitting tissue of mature styles in a region where incompatible pollen tubes are most frequently arrested.

It is interesting to note that only one band homologous to the *Nicotiana* S_2 cDNA is observed on 'Southern' blots of S_2 genomic DNA. Presumably, any male-expressed S-molecules are either products of the same gene as the female-expressed glycoprotein, as proposed in the 'dimer hypothesis' (Lewis, 1964), or are so different from the female-expressed sequence as to not cross-hybridize with it. No convincing data has been published showing expression of the cloned S-associated sequence in pollen. This may reflect low expression or other difficulties in detection, or may indicate that the male-expressed sequence is truly different from the cloned sequence.

The molecular basis of SI in *Brassica* was first investigated by Lewis (1952) who, using techniques developed for serological testing, found antigenic differences between self-incompatible stigmas of various S-genotypes. S-linked antigens have since been detected by more sensitive immunological techniques (Hinata *et al.*, 1982) and S-specific bands have been noted on iso-electric focusing gels (Nishio and Hinata, 1978).

The substance implicated in the specification of SI in *Brassica* stigmas was shown to be a glycoprotein of ca 60KDa with an alkaline pI value. Several lines of evidence (Nasrallah *et al.*, 1985a), in parallel to the findings in *Nicotiana*, strongly suggest this molecule to be the product of the S-locus: the gene encoding it is tightly S-linked, the glycoprotein is found only in stigmatic tissue and its synthesis correlates with the onset of SI.

In order to clone one allelic form of the gene encoding the S-locus specific glycoprotein (SLSG), Nasrallah *et al.* (1985b) established a *Brassica* S_6 stigmatic cDNA library. This was screened with total mature stigma and seedling cDNA probes. Clones hybridising strongly and selectively to the stigma probe were selected. These were found, by cross-hybridization, to be homologous and the longest was sequenced. Following this, Takayama *et al.* (1987) published the first *Brassica* SLSG protein sequences and were able to align their S_8 protein sequence with the predicted protein sequence from the S_6 cDNA (after the addition of several base pairs in order to shift and extend the open reading frame of the latter). Nasrallah *et al.* (1987) published the protein sequences from cDNAs corresponding to the presumed

S_{13} and S_{14} *Brassica* alleles and the cDNA sequences for a presumed S-allele from S_{29} plants has been published by Trick and Flavell (1989). A cDNA sequence for the S_{22} SLSG has been published by Lalonde *et al.* (1989). The cloned cDNAs showed the SLSG glycoproteins to be of ca. 75% overall homology when considered in pairs, with a particularly variable central region and C-terminus.

A number of potential N-glycosylation sites were shown to be present in the central and N-terminal portions of the S-linked protein cores. The structure of the oligosaccharide moieties conjugated with the protein has been determined from the S_8 glycoprotein (Takayama *et al.*, 1986). Two types of oligosaccharide are present, the predominant type being of Mr = 1391 Da. Not all potential N-glycosylation sites are believed to be glycosylated: Takayama *et al.* (1987) estimated six out of nine sites to be occupied in the S_8 SLSG. In particular, a glycosylation site conserved in all known SLSG sequences within a cluster of eleven conserved cysteine residues is thought to remain unoccupied; presumably an oligosaccharide moiety in this region of the molecule would cause a high degree of stereochemical hindrance to the several cysteine bridges which are hypothesized to form. (Nasrallah et al., 1987).

The pattern of potential N-glycosylation sites is not identical for any pair of SLSGs for which the sequence is available. The hypothesis that glycosylation pattern is responsible for specificity in the recognition reaction therefore remains viable. The experiments of Sarkar *et al.* (1988) (referred to earlier) show glycosylation of some molecule to be required for SI in *Brassica*, though of course

the identity of this molecule could not be confirmed and it was not established that the glycosyl moieties were necessary specifically as recognition components rather than as some other essential component. The highly variable central region of the SLSG molecule may have a role in specificity, though no firm evidence exists to support either case; clearly both glycosylation pattern and sequence variability may play a part in specificity determination.

The genomic organisation of the SLSG gene has been investigated (Nasrallah *et al.*, 1988). A more complex genomic banding pattern is observed on 'Southern' blots than is the case with *Nicotiana* S-associated glycoprotein genes. Nasrallah *et al.* (1988) suggest the SLSG message to be the product of a single active gene and further state that a number of apparent pseudogenes are present in the genomic region of the S-locus. *In situ* hybridization studies show the SLSG message to be present only in the stigmatic papillar cells, reaching a maximum concentration at one day before anthesis (as the buds become self-incompatible). No published evidence exists to indicate the transcription of an SLSG-homologous gene in anther tissue.

A further type of stigma-specific glycoprotein has been discovered in *Brassica* (Isogai *et al.*, 1988, Lalonde *et al.*, 1989, Trick and Flavell, 1989). This bears ca. 50% homology to the SLSG molecules, but appears to be invariant between lines carrying different S-alleles. Lalonde *et al.* (1989) demonstrate that this S-related molecule (termed the SLR glycoprotein or NS glycoprotein) is not closely S-linked as segregation of SLR-homolgous restriction fragment-length polymorphisms from S-specificities is noted in an F_2

generation. The significance of this glycoprotein is totally unknown.

The mechanisms and origins of SI in angiosperms clearly require much further investigation. The principal questions which may be addressed by the use of molecular biological techniques include the following:-

1. What is the pollen recognition component in SI?
2. How do the components of SI interact in recognition? (This may also give clues as to the functioning of the next step in signal transduction which leads to self-pollen rejection.)
3. What is the basis of S-allele diversity and (relatedly) how are new S-allele specificities generated?
4. Along what routes did the present-day systems of SI evolve? (Homologies between cloned S-gene sequences are expected to answer this question.)

1.5 The Aims of the Present Work

The approach taken in this work was to clone a cDNA representing one allele of the SLSG gene in *Brassica oleracea*, to determine its sequence and to investigate its genomic organisation and expression by means of blot hybridization experiments. DNA sequence data for a novel S-allele was considered desirable because, at the outset, only one SLSG cDNA sequence was published and clearly comparison between such sequences, (whose role in specificity depends on mutual dissimilarity), is essential to an understanding of SI.

The particular S-allele chosen was S_5, which is the most recessive of all *Brassica* S-alleles (Ockendon, 1975) and is used in plant breeding for the production of F_1 hybrid varieties.

The further aims of this project were formulated as results emerged. Some effort was made in attempting to find expression of an SLSG-homologous sequence in anther tissue. This was based on the hypothesis that male and female-expressed S-gene sequences may be closely related. An attempt was made to select cDNAs, apart from the SLSG cDNA, which were also stigma specifically expressed. The rationale for doing this was based on the possibility that a number of stigmatically-expressed sequences may be involved in the mechanism of SI, even though specificity is governed by a single genetic locus.

The cloning of an SLSG-like cDNA made possible the investigation of homology with the presumed *Brassica* S-gene in other species. This line of experimentation was aimed at the elucidation of the level of evolutionary divergence of the *Brassica oleracea* SI system from other sporophytic SI systems.

CHAPER TWO

MATERIALS

2 MATERIALS

2.1. Plant Material

(Gene bank accession numbers are given in parentheses.)

Seed material of *Brassica oleracea* var *gemmifera* S_5 (DJ6005), *B. oleracea* var. *alboglabra* S_5 (DJ4146), *B. carinata* (002485), *B. juncea* (002487) and *B. nigra* (002490) was obtained courtesy of Dr. D. Ockendon, IHR, Wellesbourne, Warwickshire. In addition, frozen leaf material of *B. oleracea* inbred lines containing S-alleles: S_9 (DJ8153), S_{16} (DJ6050), S_{18} (DJ8018), S_{20} (DJ4734), S_{39} (DJ4763) and S_{58} (DJ6029) was obtained courtesy of Drs. D. Ockendon and G. King, IHR, Wellesbourne, Warwickshire.

Seed material of : *Brassica rapa* (0020747), *Crambe maritima* (0034327), *Iberis amara* (000332) and *Senecio viscosus* (00020655) was obtained courtesy of Royal Botanic Gardens (Kew), Wakehurst Place, Haywards Heath, Sussex.

Seeds of *Raphanus sativus* cv. Long White Icicle were obtained from Mr. Fothergills Seeds, Kentford, Suffolk.

Seeds of *Hesperis matrionalis* and plants of *Reseda lutea* were obtained courtesy of Dr. P. Gates, Department of Biological Sciences, University of Durham.

2.2 Bacterial Strains

Frozen competent cells of *Escherishia coli* DH5α of library grade cloning efficiency were obtained from BRL Ltd., P.O. Box 35, Trident House, Renfrew Road, Paisley, PA3 4EF. *E. coli* JM101 (Yanisch-Perron *et al.* 1985) was available within the Department of Biological Sciences, University of Durham.

2.3 DNA

pUC18, pBR322 and pHA1 plasmid preparations were obtained courtesy of Dr. R. Croy, Department of Biological Sciences, University of Durham (pHA1 was originally obtained from Dr, R, Cuellar, Plant breeding Institute, Cambridge). pAT1 was generously donated by Mr. A. Thompson and *Arabidopsis thaliana* genomic DNA was generously donated by Mr. S. Yaish, both of the Department of Biological Sciences, University of Durham. M13mp18, M13mp19 and bacteriophage λ DNA were obtained from BCL Ltd., Lewes, East Sussex.

Oligonucleotide A was generously donated by Dr. C. Franklin of the School of Biological Sciences, University of Birmingham. It is a 30mer sequence representing bases 100 to 130 of the coding strand of the S_6 SLSG cDNA from *Brassica oleracea* (Nasrallah *et al.*, 1985). Oligonucleotide B was synthesised by Mr. J. Gilroy of the Department of Biological Sciences, University of Durham. It is a 30mer corresponding to bases 1096 to 1126 of the coding strand of the above mentioned sequence.

Synthetic EcoRI DNA linkers were purchased from Pharmacia Fine Chemicals, Midsummer Boulevard, Milton Keynes, MK9 3HP.

2.4 Chemicals and Media Constituents Etc.

All chemicals were purchased from BDH Ltd., Poole, Dorset unless otherwise indicated. All chemicals were of analytical grade or the best available.

Acrylamide, acridine orange, ATP, ampicillin, bovine serum albumin, ethidium bromide, glyoxal, HEPES, herring sperm DNA, IPTG, laurylosarcosine, MES, N,N methylene-bis-

acrylamide, polyvinylpyrrolidone, pronase, RNase A and tRNA were obtained from Sigma Chemical Co. Poole, Dorset.

Restriction endonucleases and DNA modifying enzymes were obtained from Northumbrian Biologicals Ltd., Cramlington, Northumberland and BCL Ltd., Lewes, East Sussex. cDNA synthesis and M13 sequencing kits were also obtained from BCL Ltd.

Radiolabelled compounds and 'Hybond-N' membranes were obtained from Amersham International p.l.c., Amersham, Oxon.

Nitrocellulose membranes were obtained from Scliecher and Schuell Gmbl, Dassel, F.R.G.

GFC filters and 3MM chromatography paper were obtained from Whatman Ltd., Maidstone, Kent.

PPO and POPOP were obtained from Koch Light Ltd, Colnebrook, Berkshire. EcoScint A was obtained from National Diagnostics, 1013-1017 Kennedy Boulevard, Manville, New Jersey, USA.

Bacto-agar was obtained from Difco, Detroit, Michigan, USA. Bactotryptone was from Oxoid Ltd., Basingstoke, Hants. Yeast extract was obtained from Biolife Srl., Milan, Italy

Ficoll 400 and sepharose CL-4B were obtained from Pharmacia Fine Chemicals, Midsummer Boulevard, Milton Keynes, MK9 3HP.

CHAPTER THREE

METHODS

3 METHODS

Recipes for solutions and media used in the following methods may be obtained from the source references quoted in the text, or, where no reference is given, from Maniatis *et al.* (1982).

3.1 Treatment of Plant Material

Plant material was grown from seed in peat-based potting compost after a period of pre-imbibition. *B. oleracea* var. *gemmifera* plants were induced to flower by long photoperiods (greater than twelve hours) following vernalisation for a period of two months. The latter was accomplished by leaving plants outside during sufficiently cold periods, or by transferring plants to a 4^0C room equipped with artificial lighting on a twelve hour on/twelve hour off cycle. *B. oleracea* var. *albolglabra* plants were induced to flower under long day conditions alone as they have no vernalisation requirement.

Plant tissues were harvested for DNA and RNA extraction directly into liquid nitrogen. Seed stocks of self-incompatible inbred lines were maintained by pollination of immature flower buds.

3.2 Microscopic Technique

Microscopic examination of plant material was conducted with a Nikon Diaphot microscope using fluorescence optics. Plants were verified as self-incompatible prior to harvesting material for use in cloning or hybridization experiments by self-pollination of open flowers and

examination of pistils stained with decolourised aniline blue (1% $^{w}/_{v}$ in 0.1M K_3PO_4) using a blue filter. The presence of distorted pollen tubes failing to penetrate the stigmatic surface was taken as evidence of self-incompatibility, although pollen from lines containing strong S-alleles may fail to germinate following self-pollination.

Anther sections were stained in ANS(0.1% $^{w}/_{v}$ in McIlvaine's citrate/phosphate buffer pH4.0) and examined by fluoresence microscopy using a blue excitation filter.

3.3 General Manipulations of Nucleic Acids

The following standard techniques for the processing of nucleic acids were used as described by Maniatis et al. (1982): restrictions and ligations of DNA; 5'-dephosphorylation of DNA using calf intestinal phosphatase; 5'-phosphorylation of DNA using T4 polynucleotide kinase; 'polishing' of DNA using E. coli DNA polymerase I; phenol extraction; chloroform extraction; ethanol precipitation; agarose and polyacrylamide gel electrophoresis.

Specific DNA restriction fragments were isolated by agarose gel electrophoresis followed by electrophoretic elution of the DNA from excised gel slices. Elution was performed using a 'minigel' apparatus, the gel slices being contained within closed dialysis membranes. Samples were eluted into small volumes of 1 x TBE buffer prior to subsequent processing by phenol and chloroform extraction and ethanol precipitation.

3.4. Quantification of Nucleic Acids

3.4.1 Spectrophotometric Determination of Nucleic Acid Concentration

Absorbances of DNA and RNA solutions were measured on a Pye Unicam SP-150 UV/Visible Spectrophotometer. Calculations of concentration and assessments of nucleic acid purity were based on extinction coefficients at 260 nm and 280 nm given for DNA and RNA in Maniatis *et al.* (1982).

3.4.2 Spectrofluorimetric Determination of DNA Concentration

Concentrations of small samples of double stranded DNA were measured using a Baird Atomic Fluoripoint Fluorimeter by the method of Thomas and Farquar (1978). This is based on the fluorescence of DNA with diaminobenzoic acid at an excitation wavelength of 405 nm.

3.4.3 Quantification of Radiolabelled DNA

The mass of tritium-labelled DNA produced in cDNA synthesis reactions was calculated from measurement of radioactivity on a Packard Tri Carb Liquid Scintillation Analyser. Incorporated radioactivity was measured following trichloracetic acid precipitation of DNA using sonicated herring sperm DNA as a co-precipitator (Maniatis *et al.*, 1982). This method was also used in order to assess the percentage of incorporation of α-[^{32}P]-dCTP into DNA probes following labelling by random priming.

3.5 Transformation of *E. coli*

3.5.1 Transformation for cDNA Cloning

Recombinant pUC18 vector DNA (Yanisch-Perron *et al.*, 1985) was used to transform *E. coli* strain DH5-α library grade, commercially prepared competent cells in accordance with the suppliers instructions.

3.5.2 Transformation for the Subcloning of DNA Fragments

For subcloning purposes, cells were made competent by the method of Hanahan (1985) which allows cells to be stored frozen in aliquots. Subcloning of DNA fragments in M13 mp18 and M13 mp19 (Yanisch-Perron *et al.*, 1985) vectors for the production of single stranded DNA molecules for sequencing purposes was achieved by transformation of *E. coli* strain JM101. This was brought about by incubation of 200 μl aliquots of competent cells with approximately 0.1 μg recombinant M13 replicative form DNA on ice for 30 mins., incubation at 42^{0}C for 90 s and plating out by the top-layer agar method (Hanahan, 1985), incorporating: IPTG (0.3 mM), X-Gal (50μl, 2% $^{w}/_{v}$ in dimethyl formamide) and 200μl of an exponentially growing culture of *E. coli* strain JM101 into the top-layer agar.

3.6 Nucleic Acid Extraction

3.6.1 Extraction of Plant DNA

Genomic DNA was extracted from leaf tissue by the method of Graham (1978). A single density gradient centrifugation in caesium chloride was employed.

3.6.2 Extraction of RNA

Extraction of RNA from plant tissues was achieved using a method based on initial homogenisation of tissue in a buffer containing guanidinium hydrochloride and β-mercaptoethanol (Logeman *et al.*, 1987). The tissues were first ground to powders under liquid nitrogen using a mortar and pestle.

3.6.3 Preparation of Plasmid DNA

Plasmid DNA was extracted from *E. coli* cultures by the alkaline lysis method of Birnboim and Doly (1979) as modified by Ish-Horowicz (Maniatis *et al.*, 1982).

3.6.4 Preparation of Single-Stranded M13 Bacteriophage DNA

Recombinant M13 bacteriophage DNA was purified from 2 ml cultures of transformed *E. coli* strain JM101 grown in 2xYT broth. Cells and cell debris were removed by centrifugation (10,000 g, 10 min.). Phage particles were then precipitated by the addition of one-fifth volume of a solution containing PEG 6000 (20% $^{w}/_{v}$) and NaCl (2.5 M) and incubation at room temperature for 30 min. followed by recentrifugation (10,000 g, 10 min.). Single stranded DNA was purified from the phage by resuspension in 1xTE buffer followed by phenol and chloroform extraction and ethanol precipitation. Samples were analysed on 0.7% agarose gels using non-recombinant M13 mp18 single stranded DNA as a size marker.

The orientation of inserts cloned into single restriction sites in M13 vectors was determined as described by Messing (1983).

3.7 Polyadenylated RNA Selection

Two separate methods were utilized for the purification of polyadenylated RNA from total RNA samples. Initially, a column-based method of affinity chromatography on oligo-dT cellulose (Aviv and Leder, 1972) was used. The column was constructed from a 1 ml sterile, disposable syringe packed with 0.1 g of the matrix. This was used for samples of up to 200 mg of total RNA.

A batch method was subsequently designed utilising the same matrix and buffers as the column method. In this method, RNA samples were processed in 1.5 ml eppendorf tubes containing 0.1 g oligo-dT cellulose. Samples were equilibrated with the oligo-dT cellulose in 1 ml of loading buffer and washed five times in this solution, centrifuging after each wash (6,000 rpm on an MSE microcentaur microcentrifuge, one min.). Polyadenylated RNA was removed by washing the matrix four times, each with 0.25 ml of wash buffer. The washes were pooled and RNA was ethanol precipitated using 10μg mussel glycogen as a co-precipitator. (The latter was found by experimentation not to affect the efficiency of cDNA synthesis reactions.)

3.8 cDNA Synthesis and Cloning

cDNA was synthesised using a kit in accordance with the manufacturers instructions. The method of synthesis employed was essentially that of Gubler and Hoffman (1983) with the addition of a T_4 DNA polymerase 'polishing' step in order to render a high proportion of the cDNA molecules blunt ended. (This method yields a high proportion of full-length cDNA clones in comparison to the older method of

cDNA synthesis which involves the use of S_1-nuclease.)

cDNA was ligated into pUC18 cloning vector via the addition of synthetic EcoR1 linkers. Linkers were phosphorylated by the action of T_4 polynucleotide kinase using a pulse of γ-[^{32}P]-ATP and a 'cold chase' of unlabelled ATP. Linkers were ligated to cDNA and restricted by prolonged treatment with EcoRI. The results of ligation of linkers and subsequent restriction was analysed by electrophoresis of appropriate samples on 10% polyacrylamide gels and autoradiography of these. Linker monomers were separated from linkered cDNA by gel filtration (Watson and Jackson, 1985). The linkered cDNA fractions were pooled and concentrated by ethanol precipitation prior to ligation into the cloning vector and transformation for library construction as previously described.

Libraries were prepared by two methods. A library of 2000 recombinant clones was prepared and stored in microtitre plates (one clone per well) as described by Mason and Williams (1985). A second library of approximately 20,000 recombinant clones was prepared for oligonucleotide screening as described by Woods (1984). In this case, clones were stored on nitrocellulose filters overlaid on glycerol-containing agar plates. Screening was accomplished on duplicate filters produced by contact with the originals.

3.9 Hybridization Experiments

3.9.1 DNA Dot Blotting

DNA dot blots were prepared for hybridization experiments by immobilization on nitrocellulose membranes as described by Marzluff and Huang (1984).

3.9.2 Southern Transfer of DNA

DNA samples were transferred to nitrocellulose or nylon filters from neutral agarose gels essentially as described by Southern (1975). Gels were incubated in HCl (0.1 M) for fifteen min., denaturing solution for one hour and neutralizing solution for one hour. Transfer was mediated by a solution of 20xSSC, the membrane having been previously equilibrated in this. DNA samples were immobilized by baking at 80^{0}C for one hour, in the case of nitrocellulose filters and by appropriate UV exposure (as directed by the manufacturers) in the case of nylon filters.

3.9.3 Northern Transfer of RNA

Northern transfer was carried out as described by Thomas (1980) following denaturation of RNA samples in dimethylsulphoxide and glyoxal as described by McMaster and Carmichael (1977). Glyoxalated DNA size markers were included on the gels used for Northern blotting and appropriate portions of the gels were removed to enable the visualization of these and of ribosomal RNA bands (for assessment of RNA integrity). These gel portions were stained in acridine orange and examined by UV transillumination (McMaster and Carmichael, 1977).

3.9.4 Preparation of Bacterial Colonies for Hybridization Experiments

Bacterial colonies were processed for cDNA library screening on nitrocellulose filters as described by Mason and Williams (1985).

3.9.5 Hybridization of Oligonucleotide Probes to Bacterial Colonies and Southern Blots

Oligonucleotide hybridizations were carried out at low stringency as they involved the formation of imperfectly matched hybrids. 30-mer oligonucleotides were end labelled by the action of T_4 polynucleotide kinase using equimolar amounts of γ-[^{32}P]-ATP (3000 Ci/mmol). Pre-hybridizations were performed in a shaking water bath at 35^{0}C for four hours in solutions of: 5xSSC, 5xDenhardt's reagent, SDS (0.1% $^{w}/_{v}$), sonicated, denatured herring sperm DNA (100 μg/ml) and tetrasodium pyrophosphate (0.05% $^{w}/_{v}$). Hybridizations were carried out under similar conditions, but for sixteen hours with the addition of the denatured DNA probes. Filters were washed in solutions containing 5xSSC and SDS (0.1% $^{w}/_{v}$), initially at 40^{0}C and then at 5^{0}C increments. Autoradiography was performed after each wash until signals arising from hybridizing sequences could be detected above background radiation. Once conditions of stringency for a particular oligonucleotide and its target sequence had been established, subsequent hybridizations could be performed by washing immediately at the highest stringency possible.

3.9.6 Southern Hybridization using Polynucleotide DNA Probes

Southern hybridizations were performed using polynucleotide probes labelled using α-[^{32}P]-dCTP (400 Ci/mmol) by a 'random priming' method (Feinberg and Vogelstein, 1983). Hybridizations were carried out as for oligonucleotide hybridizations (above), but at 65^{0}C.

Washing of filters was usually performed at 65^{0}C, stringency being adjusted by the use of various salt concentrations in the final wash.

3.9.7 Northern Hybridization Using Polynucleotide DNA Probes

Northern hybridizations were performed using probes labelled as for Southern hybridizations with polynucleotide DNA probes (above). Filters were pre-hybridized for four hours at 42^{0}C in a solution containing: deionised formamide (50% $^v/_v$), 5xSSPE, 5xDenhardt's reagent, SDS (0.1% $^w/_v$), sonicated, denatured herring sperm DNA (100 μg/ml) and tetrasodium pyrophosphate (0.05% $^w/_v$). Hybridizations were carried out as for pre-hybridization, but for sixteen hours and with the addition of the denatured DNA probe. Washing of filters was performed as for Southern hybridizations, but at 60^{0}C.

3.10 DNA Sequence Analysis

DNA clones were sequenced by the Sanger dideoxy chain-termination method (Sanger *et al.*, 1980) using both an M13 sequencing kit and, latterly, an Applied Biosystems 370 A DNA Sequencer. Certain clones were restriction mapped and their restriction fragments appropriately sub-cloned in M13 mp18 and M13 mp19 vectors such that DNA templates were produced from which the entire sequence of both strands of the DNA could be deduced. Other clones were sub-cloned such that they could be sequenced from one strand only.

Homology with predicted amino acid sequences derived from sequenced cDNA molecules was investigated, where necessary, by the use of the Protein Sequence Database of the Protein Identification Resource (PIR) compiled by the National Biomedical Research Foundation, Georgetown University Medical Centre, 3900 Reservoir Road NW, Washington DC, 20007.

CHAPTER FOUR

RESULTS AND DISCUSSION

4 RESULTS AND DISCUSSION

4.1 cDNA Library Construction and Screening

The initial objective of this project was to clone and select cDNA representatives of genes potentially involved in SI from a homozygous S_5 line of *Brassica oleracea* var. *gemmifera* plants. Selection of a cDNA putatively encoding the S_5 allele of the SLSG gene, which is implicated in encoding specificity of SI in stigmas, was performed by oligonucleotide screening at low stringency using probes derived from the published S_6 SLSG sequence (Nasrallah *et al.*, 1985b). Selection of other sequences potentially involved in SI and other intercellular signalling responses in stigmas was performed by differential screening of a stigma cDNA library using stigma and leaf total cDNA probes. The libraries used for this work were constructed using *B. oleracea* var. *gemmifera* S_5 stigma tissue harvested one day prior to anthesis. Figure 1A shows an autoradiograph of a polyacrylamide gel electrophoresis analysis of the linkered cDNA used in library construction prior to ligation into the cloning vector.

4.1.1 Oligonucleotide Screening of cDNA Libraries

Initial oligonucleotide screening was performed on a 2000 clone cDNA library using a 30-mer oligonucleotide probe derived from residues 100 to 130 of the coding strand of the S_6 SLSG cDNA sequence (Nasrallah *et al.*, 1985b). This probe (designated oligonucleotide A) hybridized strongly to one clone at a stringency equivalent to washing at 45°C in 5xSSC. This clone (designated 3D2) was fully sequenced.

FIGURE 1 Construction and screening of *Brassica* stigma cDNA libraries.

A
Autoradiograph of a 10% polyacrylamide gel following electrophoresis of : track 1, *Brassica oleracea* var. *gemmifera* S_5 stigma cDNA after ligation to end-labelled EcoRI linkers; track 2, linkered cDNA (as in track 1) following restriction with EcoRI.

B
Secondary screening of colony hybridizations of selected cDNA clones from a *B. oleracea* var.*gemmifera* S_5 cDNA library to radiolabelled probes of: (i) total stigma cDNA; (ii) total leaf cDNA. Hybridizations were washed at high stringency (0.1 x SSC, 65°C). Leaf-expressed clones are included as controls.

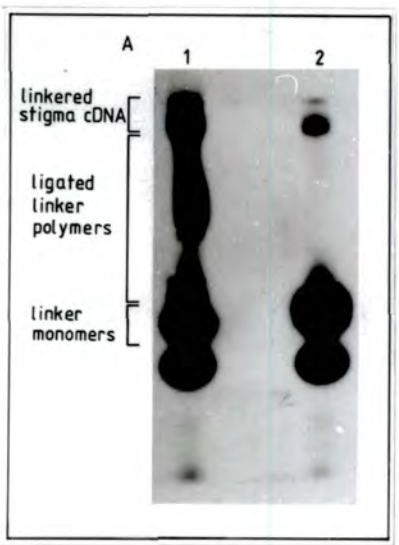
C
Autoradiographs of cross-hybridizations of selected cDNA clones. Colony hybridization was performed using selected cDNAs as probes in order to divide clones into cross-hybridizing groups. The probes used are indicated underneath each autoradiograph. Twelve cross-hybridizing groups (i to xii) are demonstrated. The hybridizations were washed at high stringency (0.1 x SSC, 65°C).

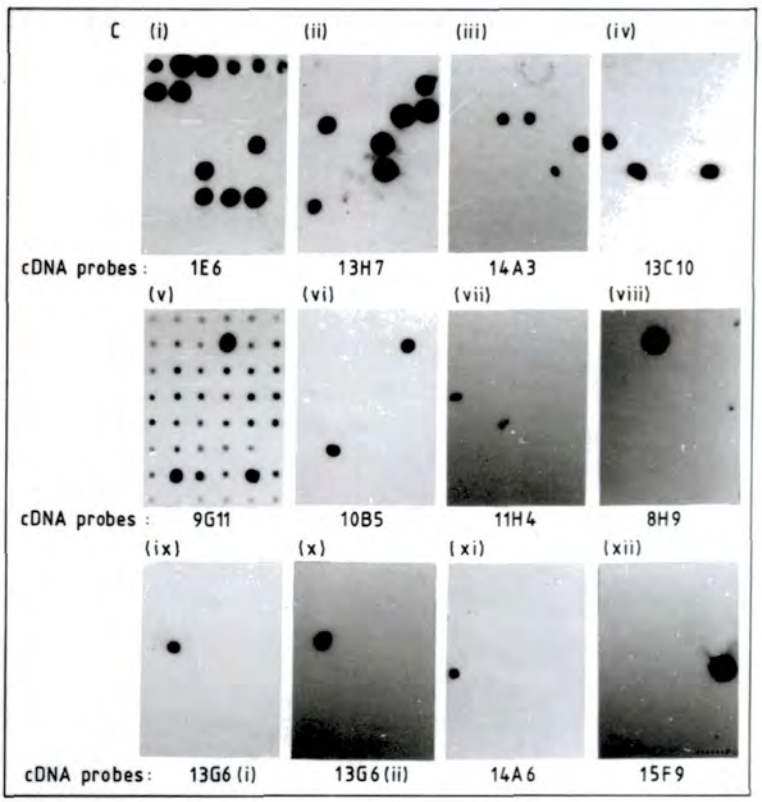
D
Autoradiographs of hybridizations of a total stigma cDNA probe from *B. oleracea* var.*gemmifera* S_5 plants to DNA dot blots of selected cDNA clones; one from each class identified in Figure 1C (above). The cDNA dot blots were: dot 1, 1E6; dot 2, 8H9; dot 3, 9G11; dot 4, 10B5; dot 5, 11H4; dot 6, 13C10; dot 7, 13G6(i); dot 8, 13G6(ii); dot 9, 14A4; dot 10, 14A6; dot 11, 15C1; dot 12, 15F9. The hybridization was washed at high stringency (0.1 x SSC, 65°C).

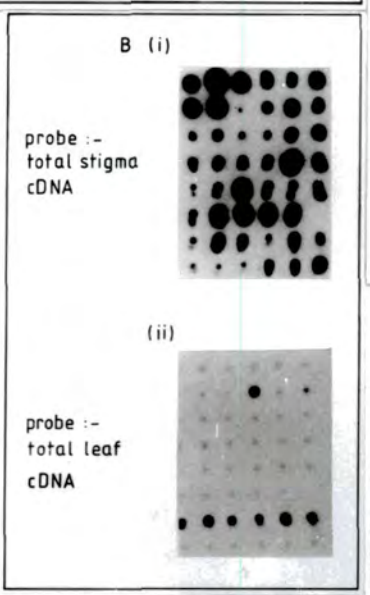
E
Autoradiograph of a hybridization of a probe of the 9.0Kb HindIII fragment from plasmid pHA1 (Cuellar, 1982), which encodes all classes of nuclear ribosomal genes from *Pisum sativum*, to dot blots of selected cDNA clones as for Figure 1D (above).

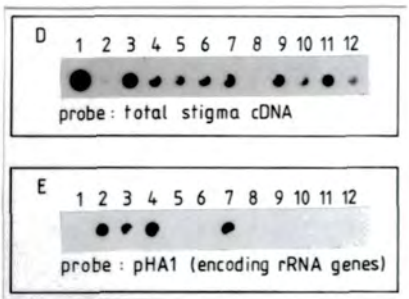
F
Autoradiograph of a 10% polyacrylamide gel following electrophoresis of end-labelled oligonucleotide B: a 30-mer corresponding to bases +1096 to +1126 of the *B.oleracea* S_6 SLSG allele (Nasrallah *et al.*, 1985b). This was used as a probe in cDNA library screening.

G
Screening of a 20 000 clone cDNA library prepared from *B. oleracea* var. *gemmifera* S_5 stigma cDNA using oligonucleotide B (see Figure 1F, above). The library was screened on 20 duplicate nitrocellulose filters, one pair of which are illustrated and contain a single hybridizing clone.

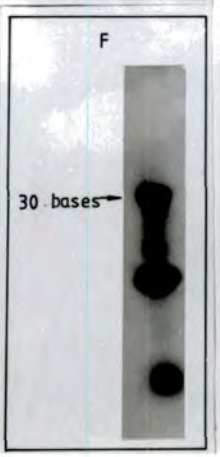

A
1
2
linkered stigma cDNA
ligated linker polymers
linker monomers


C
(i)
(ii)
(iii)
(iv)
cDNA probes : 1E6 13H7 14A3 13C10
(v)
(vi)
(vii)
(viii)
cDNA probes : 9G11 10B5 11H4 8H9
(ix)
(x)
(xi)
(xii)
cDNA probes : 13G6 (i) 13G6 (ii) 14A6 15F9


B (i)
probe :- total stigma cDNA
(ii)
probe :- total leaf cDNA


D
1 2 3 4 5 6 7 8 9 10 11 12
probe : total stigma cDNA

E
1 2 3 4 5 6 7 8 9 10 11 12
probe : pHA1 (encoding rRNA genes)


F
30 bases

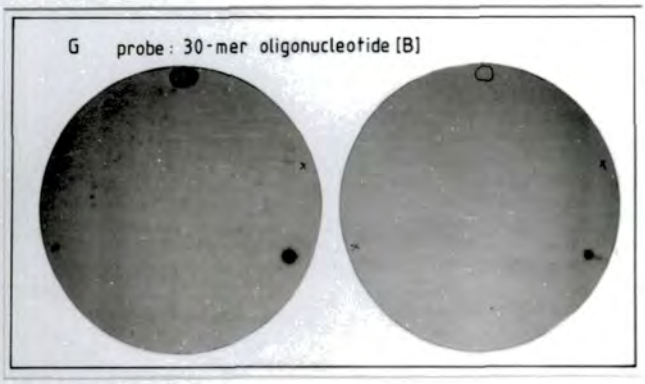

G
probe : 30-mer oligonucleotide [B]

Figure 2A shows an autoradiograph of a Southern blot hybridization of the plasmid prepared from clone 3D2 and of two others from the same cDNA library. The lower band in tracks 1 to 3 represents hybridization of oligonucleotide A to the excised cDNA, whilst the fainter upper band represents a small quantity of incompletely restricted plasmid which contains the cDNA. 3D2 contains a partial cDNA of 447 bp in which the best match of sequence with oligonucleotide A which can be found by dot matrix comparison is 21 in 30 contiguous base pairs. The cDNA has no significant extensive homology with the SLSG gene and no identification of it can be made using the available data base (see Section 3.10).

A second oligonucleotide, representing bases 1096 to 1126 of the S_6 SLSG cDNA sequence (Nasrallah *et al.*, 1985b) was synthesized and used to screen the cDNA library as for oligonucleotide A. This oligonucleotide (designated oligonucleotide B) represents a region which is conserved between the predicted amino acid sequence of the S_6 SLSG cDNA and the protein sequence data from the *B. campestris* S_8, S_9 and S_{12} alleles (Takayama *et al.*, 1987). No clone from the cDNA library was found to hybridize to oligonucleotide B at very low stringency.

A library of ca. 20,000 cDNA clones was constructed and screened with oligonucleotide B. Three clones hybridized strongly with the probe. Figure 1G shows an autoradiograph of a PAGE analysis of the labelled oligonucleotide. Figure 1F shows the use of this oligonucleotide in probing the cDNA library. One of the three hybridizing clones (in duplicate) is illustrated. These three clones contain cDNAs of some

FIGURE 2

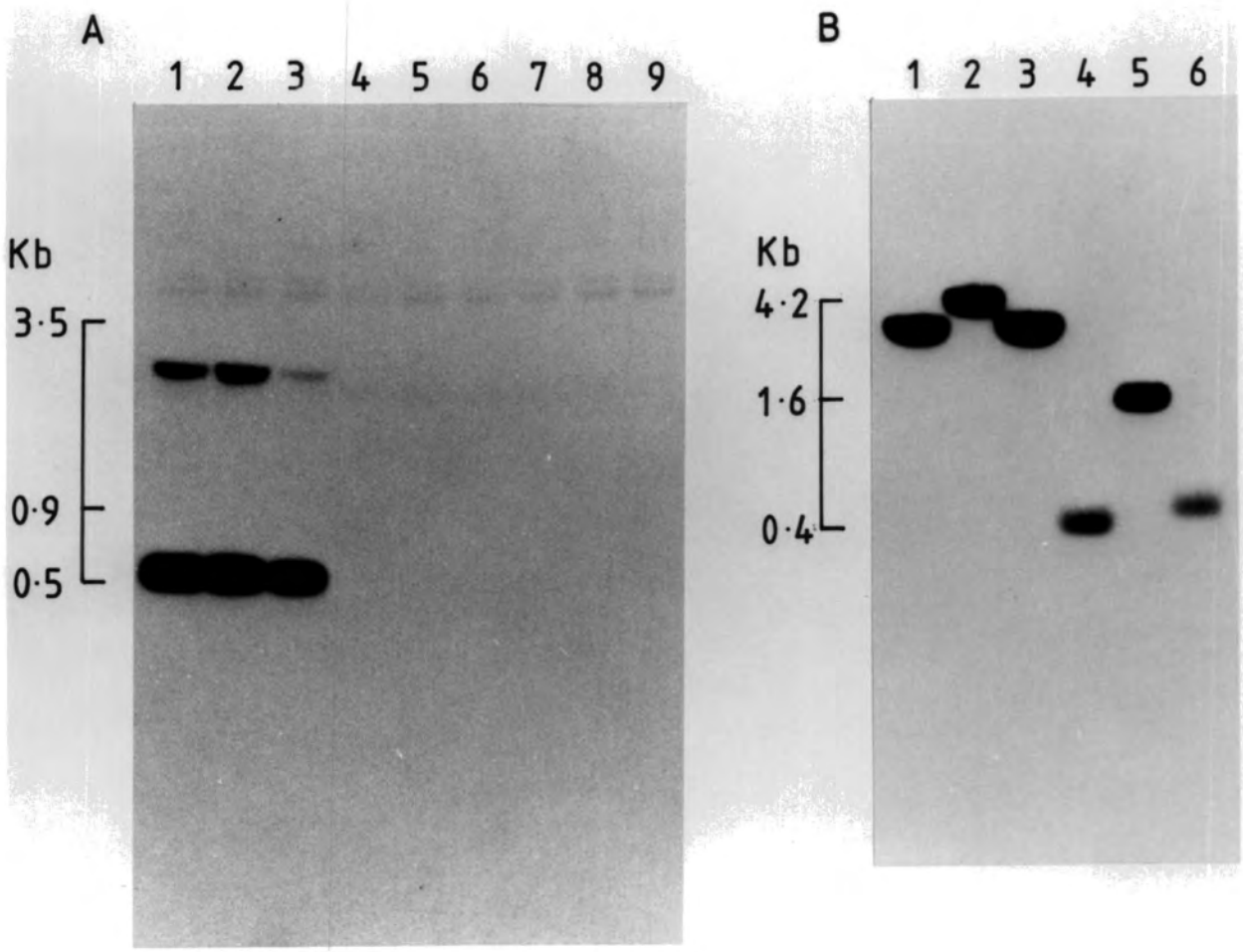
Southern hybridizations of cDNA clones with end-labelled oligonucleotide probes. Restricted plasmids containing cDNAs were electrophoresed on 0.7% agarose gels prior to Southern transfer.

A

Autoradiograph of a Southern hybridization of replicates of plasmids containing inserted cDNAs: 3D2, tracks 1, 2 and 3; 7B5, tracks 4, 5 and 6; 16E11, tracks 7, 8 and 9. The plasmids were restricted with EcoRI and electrophoresed on a 0.7% agarose gel prior to Southern transfer. The blot was probed with oligonucleotide A: a 30-mer homologous to bases +100 to +130 of the coding strand of the S_6 SLSG cDNA sequence (Nasrallah *et al.*, 1985b). Final washing of the hybridization was in 5 x SSC at 45°C.

B

Autoradiograph of a Southern hybridization of recombinant plasmids: pBS1, tracks 1 and 4; pBS2, tracks 2 and 5; pBS3, tracks 3 and 6. The plasmid DNA samples (1μg per track) were restricted with; SmaI, tracks 1, 2 and 3; and EcoRI, tracks 4, 5 and 6, and electrophoresed on a 0.7% agarose gel prior to Southern transfer. The blot was probed with oligonucleotide B: a 30-mer homologous to bases +1096 to +1126 of the coding strand of the S_6 SLSG cDNA (Nasrallah et al., (1985b). Final washing of the hybridization was in 5 x SSC at 55°C.


A
1 2 3 4 5 6 7 8 9
Kb
3·5
0·9
0·5
B
1 2 3 4 5 6
Kb
4·2
1·6
0·4

400, 1600 and 500 bp and are designated pBS1, pBS2 and pBS3, respectively. Figure 2B shows an autoradiograph of a Southern hybridization of the plasmids from the selected clones probed with oligonucleotide B. Tracks 1 to 3 represent linearised plasmids whilst in tracks 4 to 6 the cDNAs have been excised from the plasmid. Cross hybridization data (not presented) indicate the three cDNAs to be closely homologous. A full nucleotide sequence of clone pBS2 was obtained, which showed this clone to be 70.8% homologous to the S_6 SLSG sequence (Section 4-2-1).

4.1.2 Differential Screening of a cDNA Library

A stigma cDNA library of 2000 clones was screened with total stigma cDNA and total leaf cDNA probes. Thirty-five clones which hybridized strongly with stigma cDNA, but not with leaf cDNA, were selected. These clones were re-screened with the two cDNA probes, together with some clones which hybridized strongly to leaf cDNA (Figure 1B).

Plasmids from selected clones were prepared and cDNAs from these were labelled and used as probes in colony hybridizations with all of the clones selected from the cDNA library by differential screening. Figure 1C shows colony hybridizations of the selected cDNA clones to one cDNA probe from each homologous group of cDNAs. This process of cross-hybridization identified twelve groups of homologous cDNAs. The screening of cDNA dot blots from each group of homologous cDNAs with a total stigma cDNA probe (Figure 1D) showed the cDNA designated 13G6(i), but not 13G6(ii), which was cloned within the same plasmid, to be highly expressed in stigmas. Screening of dot blots of the cDNAs with a

ribosomal gene probe (Figure 1E) identified a number of groups of clones to be cDNA copies of rRNA. These were selected because the stigma total cDNA probe used (Figure 1B) was presumably contaminated with ribosomal sequences whereas the leaf total cDNA probe was not.

Of the remaining classes of clones, one was shown by sequencing to contain no putative coding region (not presented). The remaining cDNAs, which represent mRNA species, are members of four groups. The longest clone of each group was sequenced. These cDNAs were designated 1E6, 11H4, 13G6(i) and 15H11.

4.2 Analysis of cDNA pBS2: a Putative SI Sequence

Clone pBS2 was selected by oligonucleotide screening of a cDNA library using a probe representing a part of the coding sequence of the S_6 SLSG cDNA (Nasrallah *et al.*, 1985b) which, on the available evidence, was considered likely to be within a conserved region. The oligonucleotide was found to become dissociated from its target sequence during washing at between 55°C and 60°C in 5xSSC. Sequence analysis showed the region of closest homology between the cDNA and probe to contain five mismatches, the longest stretch of uninterrupted homology within this being fourteen base pairs.

4.2.1 Sequence Analysis of pBS2 cDNA

The cDNA within pBS2 was restriction-mapped and sub-cloned, initially in four fragments, into M13mp18 and M13mp19 vectors. Sequence analysis of the eight sub-clones generated (each fragment having been cloned in both

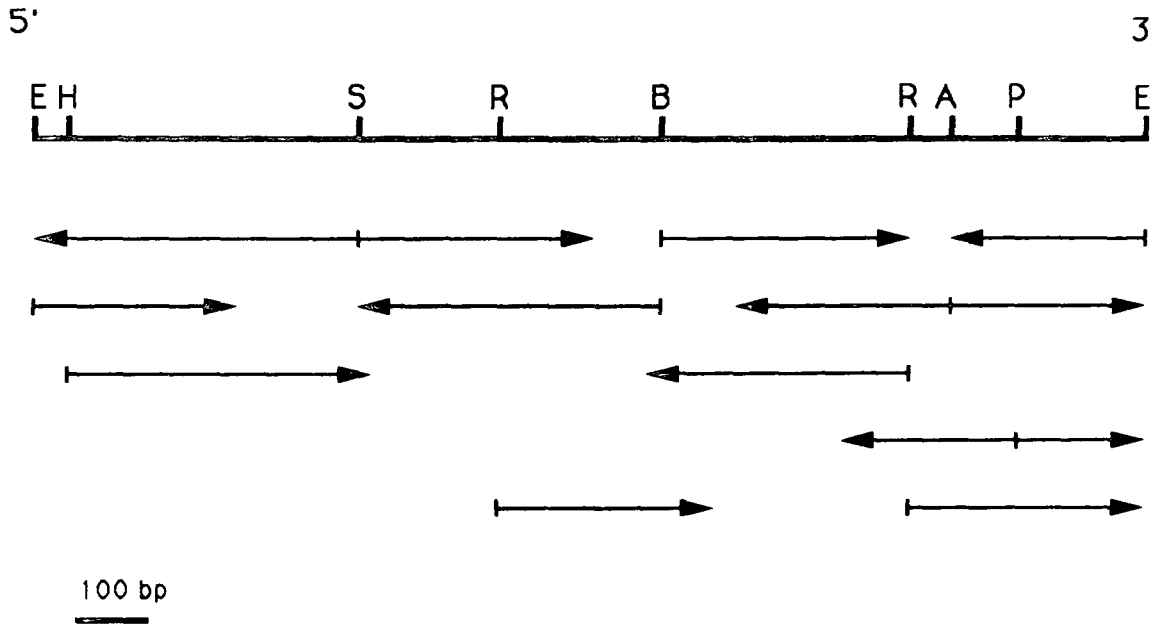
orientations) yielded a number of restriction sites from which further sub-clones of the cDNA were generated for sequencing purposes. Figure 3 shows a map of the cDNA including restriction sites which were used in the sequencing process. The nucleotide sequence data obtained from each sub-clone is represented by an arrow underneath the restriction map. The sequence of both strands of the cDNA were fully determined.

The coding strand of the pBS2 cDNA contains an open reading frame of 1284 b (427 amino acids), a 3' non-coding region of 178 b in which two sequences conforming to the polyadenylation signal motif are present, and a poly(A) tail. The cDNA has an anomalous extreme 5' end which bears a poly(dT) 'tail'. This suggests the cloned cDNA to have a short stretch of a second cDNA ligated at its 5' end. The sequence alignment (Figure 4) shows that the cDNA maintains close homology with the S_6 SLSG sequence (Nasrallah *et al.*, 1985) for some distance 5' to the N-terminus of the mature protein, though no in-frame start codon is present.

The pBS2 cDNA is 71% homologous to the S_6 SLSG cDNA sequence, whilst the predicted amino acid sequences derived from these two cDNAs are 64% homologous. The sequences of five further SLSG-like molecules are known from *Brassica oleracea* lines: S_{13} and S_{14} (Nasrallah *et al.*, 1987), S_{22} (Lalonde *et al.*, 1989) and S_{29} (Trick and Flavell, 1989) and from the *B. campestris* S_8 line (Takayama *et al.*, 1987). The pBS2 cDNA is significantly less homologous to the S_6 SLSG gene than is the BS29-2 sequence (Trick and Flavell, 1989) obtained from a *Brassica oleracea* inbred line carrying the S_{29} S-allele; the BS29-2 sequence shows 18% greater

Figure 3

Restriction and Sequencing Map of the cDNA Clone from Plasmid pBS2

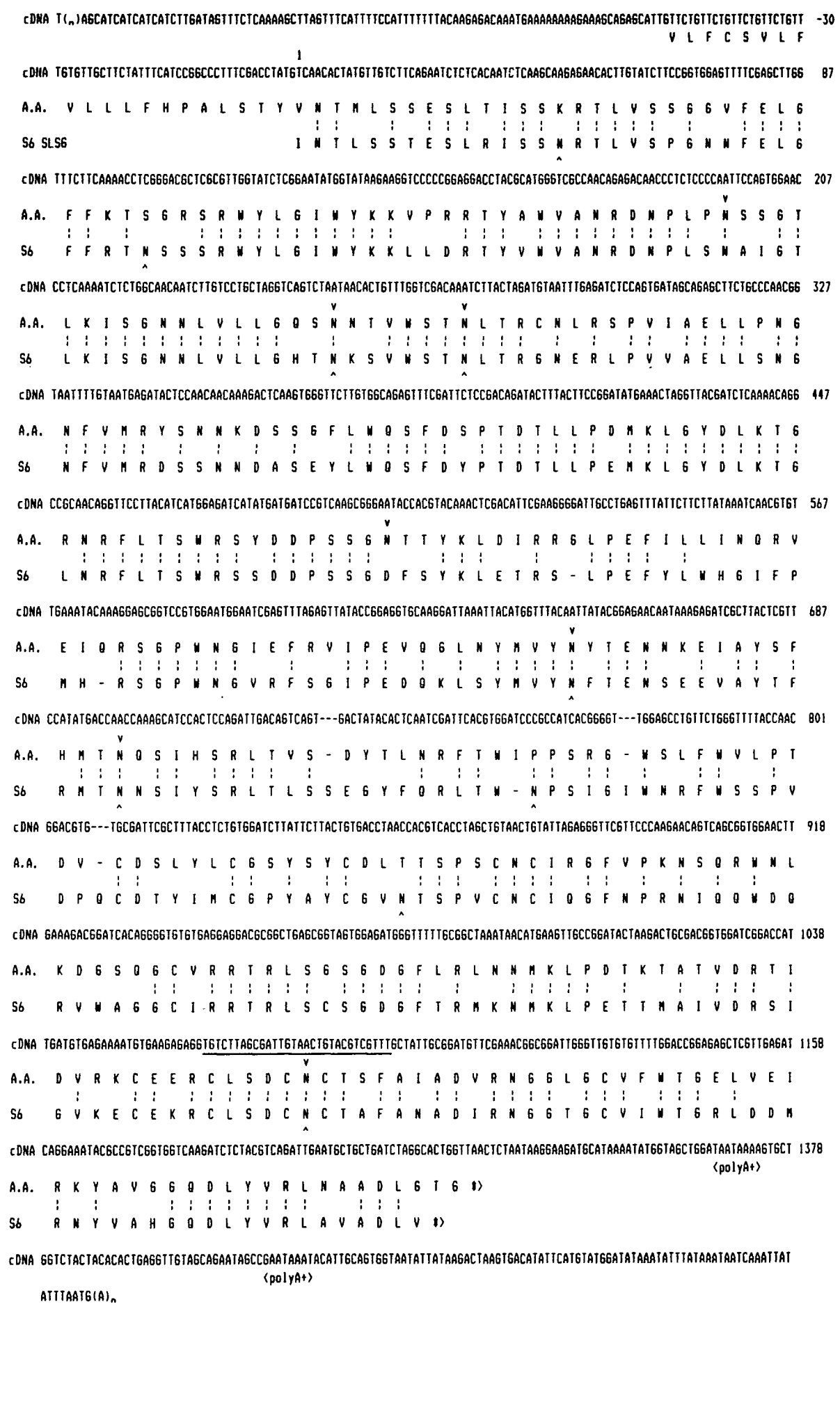


The horizontal bar represents cDNA pBS2. Restriction endonuclease sites used in subcloning for sequence analysis are indicated by letters shown above the bar according to the key shown below. The arrows represent nucleotide sequence data derived from individual recombinant M13 subclones.

E = Eco RI
H = Hind III
S = Sal I
R = Rsa I
B = Bam HI
A = Sac I
P = Hpa I

FIGURE 4

Nucleotide sequence of the pBS2 cDNA. The predicted amino acid sequence derived from the cDNA is presented immediately below it and compared with that derived from the S_6 SLSG cDNA (Nasrallah *et al.*, 1985b). Putative N-glycosylation signals are indicated by arrowheads. The region of the pBS2 cDNA putatively hybridizing to oligonucleotide probe B, derived from the S_6 SLSG cDNA sequence, is underlined. The position of the first codon of the mature protein (on the basis of homology to the S_6 SLSG molecule) is numbered 1. Putative polyadenylation signals are indicated by: <polyA+>.


cDNA T(n)A6CATCATCATCATCTTGATAGTTTCTCAAAAGCTTAGTTTCATTTTCCATTTTTTTACAAGAGACAAATGAAAAAAAAGAAAGCAGAGCATTGTTCTGTTCTGTTCTGTTCTGTT -30
V L F C S V L F
1
cDNA TGTGTTGCTTCTATTTCATCCGGCCCTTTCGACCTATGTCAACACTATGTTGTCTTCAGAATCTCTCACAATCTCAAGCAAGAGAACACTTGTATCTTCCGGTGGAGTTTTCGAGCTTGG 87
A.A. V L L L F H P A L S T Y V N T M L S S E S L T I S S K R T L V S S G G V F E L G
S6 SLSG I N T L S S T E S L R I S S N R T L V S P G N N F E L G
cDNA TTTCTTCAAAACCTCGGGACGCTCGCGTTGGTATCTCGGAATATGGTATAAGAAGGTCCCCCGGAGGACCTACGCATGGGTCGCCAACAGAGACAACCCTCTCCCCAATTCCAGTGGAAC 207
A.A. F F K T S G R S R W Y L G I W Y K K V P R R T Y A W V A N R D N P L P N S S G T
S6 F F R T N S S S R W Y L G I W Y K K L L D R T Y V W V A N R D N P L S N A I G T
cDNA CCTCAAAATCTCTGGCAACAATCTTGTCCTGCTAGGTCAGTCTAATAACACTGTTTGGTCGACAAATCTTACTAGATGTAATTTGAGATCTCCAGTGATAGCAGAGCTTCTGCCCAACGG 327
A.A. L K I S G N N L V L L G Q S N N T V W S T N L T R C N L R S P V I A E L L P N G
S6 L K I S G N N L V L L G H T N K S V W S T N L T R G N E R L P V V A E L L S N G
cDNA TAATTTTGTAATGAGATACTCCAACAACAAAGACTCAAGTGGGTTCTTGTGGCAGAGTTTCGATTCTCCGACAGATACTTTACTTCCGGATATGAAACTAGGTTACGATCTCAAAACAGG 447
A.A. N F V M R Y S N N K D S S G F L W Q S F D S P T D T L L P D M K L G Y D L K T G
S6 N F V M R D S S N N D A S E Y L W Q S F D Y P T D T L L P E M K L G Y D L K T G
cDNA CCGCAACAGGTTCCTTACATCATGGAGATCATATGATGATCCGTCAAGCGGGAATACCACGTACAAACTCGACATTCGAAGGGGATTGCCTGAGTTTATTCTTCTTATAAATCAACGTGT 567
A.A. R N R F L T S W R S Y D D P S S G N T T Y K L D I R R G L P E F I L L I N Q R V
S6 L N R F L T S W R S S D D P S S G D F S Y K L E T R S - L P E F Y L W H G I F P
cDNA TGAAATACAAAGGAGCGGTCCGTGGAATGGAATCGAGTTTAGAGTTATACCGGAGGTGCAAGGATTAAATTACATGGTTTACAATTATACGGAGAACAATAAAGAGATCGCTTACTCGTT 687
A.A. E I Q R S G P W N G I E F R V I P E V Q G L N Y M V Y N Y T E N N K E I A Y S F
S6 M H - R S G P W N G V R F S G I P E D Q K L S Y M V Y N F T E N S E E V A Y T F
cDNA CCATATGACCAACCAAAGCATCCACTCCAGATTGACAGTCAGT---GACTATACACTCAATCGATTCACGTGGATCCCGCCATCACGGGGT---TGGAGCCTGTTCTGGGTTTTACCAAC 801
A.A. H M T N Q S I H S R L T V S - D Y T L N R F T W I P P S R G - W S L F W V L P T
S6 R M T N N S I Y S R L T L S S E G Y F Q R L T W - N P S I G I W N R F W S S P V
cDNA GGACGTG---TGCGATTCGCTTTACCTCTGTGGGATCTTATTCTTACTGTGACCTAACCACGTCACCTAGCTGTAACTGTATTAGAGGGTTCGTTCCCAAGAACAGTCAGCGGTGGAACTT 918
A.A. D V - C D S L Y L C G S Y S Y C D L T T S P S C N C I R G F V P K N S Q R W N L
S6 D P Q C D T Y I M C G P Y A Y C G V N T S P V C N C I Q G F N P R N I Q Q W D Q
cDNA GAAAGACGGATCACAGGGGTGTGTGAGGAGGACGCGGCTGAGCGGTAGTGGAGATGGGTTTTTGCGGCTAAATAACATGAAGTTGCCGGATACTAAGACTGCGACGGTGGATCGGACCAT 1038
A.A. K D G S Q G C V R R T R L S G S G D G F L R L N N M K L P D T K T A T V D R T I
S6 R V W A G G C I R R T R L S C S G D G F T R M K N M K L P E T T M A I V D R S I
cDNA TGATGTGAGAAAATGTGAAGAGAGGTGTCTTAGCGATTGTAACTGTACGTCGTTTGCTATTGCGGATGTTCGAAACGGCGGATTGGGTTGTGTGTTTTGGACCGGAGAGCTCGTTGAGAT 1158
A.A. D V R K C E E R C L S D C N C T S F A I A D V R N G G L G C V F W T G E L V E I
S6 G V K E C E K R C L S D C N C T A F A N A D I R N G G T G C V I W T G R L D D M
cDNA CAGGAAATACGCCGTCGGTGGTCAAGATCTCTACGTCAGATTGAATGCTGCTGATCTAGGCACTGGTTAACTCTAATAAGGAAGATGCATAAAATATGGTAGCTGGATAATAAAAGTGCT 1378
(polyA+)
A.A. R K Y A V G G Q D L Y V R L N A A D L G T G ⇕)
S6 R N Y V A H G Q D L Y V R L A V A D L V ⇕)
cDNA GGTCTACTACACACTGAGGTTGTAGCAGAATAGCCGAATAAATACATTGCAGTGGTAATATTATAAGACTAAGTGACATATTCATGTATGGATATAAATATTTATAAATAATCAAATTAT
(polyA+)
ATTTAATG(A)n

homology to the S_6 SLSG gene than does the pBS2 cDNA at the nucleotide level.

The similarity between the translated amino acid sequence derived from pBS2 and other SLSG-like sequences is illustrated in Figures 5 and 6. The molecules show relatively variant central and C-terminal regions (the former of these residing approximately from residues 180 to 275) interspersed with regions of high sequence homology. All putative SLSG molecules have eleven entirely conserved cysteine residues towards their C-termini, though the S_5 and S_{14} forms lack one cysteine residue (at position 317) which the other sequences possess. The distribution profile of putative N-glycosylation signals shows some degree of conservation between the sequenced molecules, though this is not identical between any pair of sequences. (The putative N-glyclosylation signals are of the form Asn-X-Thr/Ser where X represents any amino acid {Wagh and Bahl, 1981}). A large number of serine and threonine residues are additionally present in the pBS2 translated sequence as potential sites for O-glycosylation.

The hydropathy plot of the protein sequence predicted from the pBS2 cDNA is shown in Figure 7. A hydrophobic N-terminal sequence which may function as a signal peptide in directing extracellular secretion of the molecule is present. The variable central region of the molecule is shown to be amphipathic, in slight contrast to the results obtained by Nasrallah *et al.* (1987) for the S_6 SLSG molecule, which show this region to be predominantly hydrophilic. These authors suggest the hydrophilicity of the central region to indicate it to be external in the

FIGURE 5

A comparison of putative SLSG amino acid sequences from the start of the mature protein. Gaps introduced into sequences to maximise homology are represented by dashes. Regions of the S_8 sequence which were not determined (Takayama *et al.*, 1987) are left blank. Boxed residues are identical, boxes being drawn only at positions where six or all of the seven sequences are conserved. References for the sequences are given in the Figure.

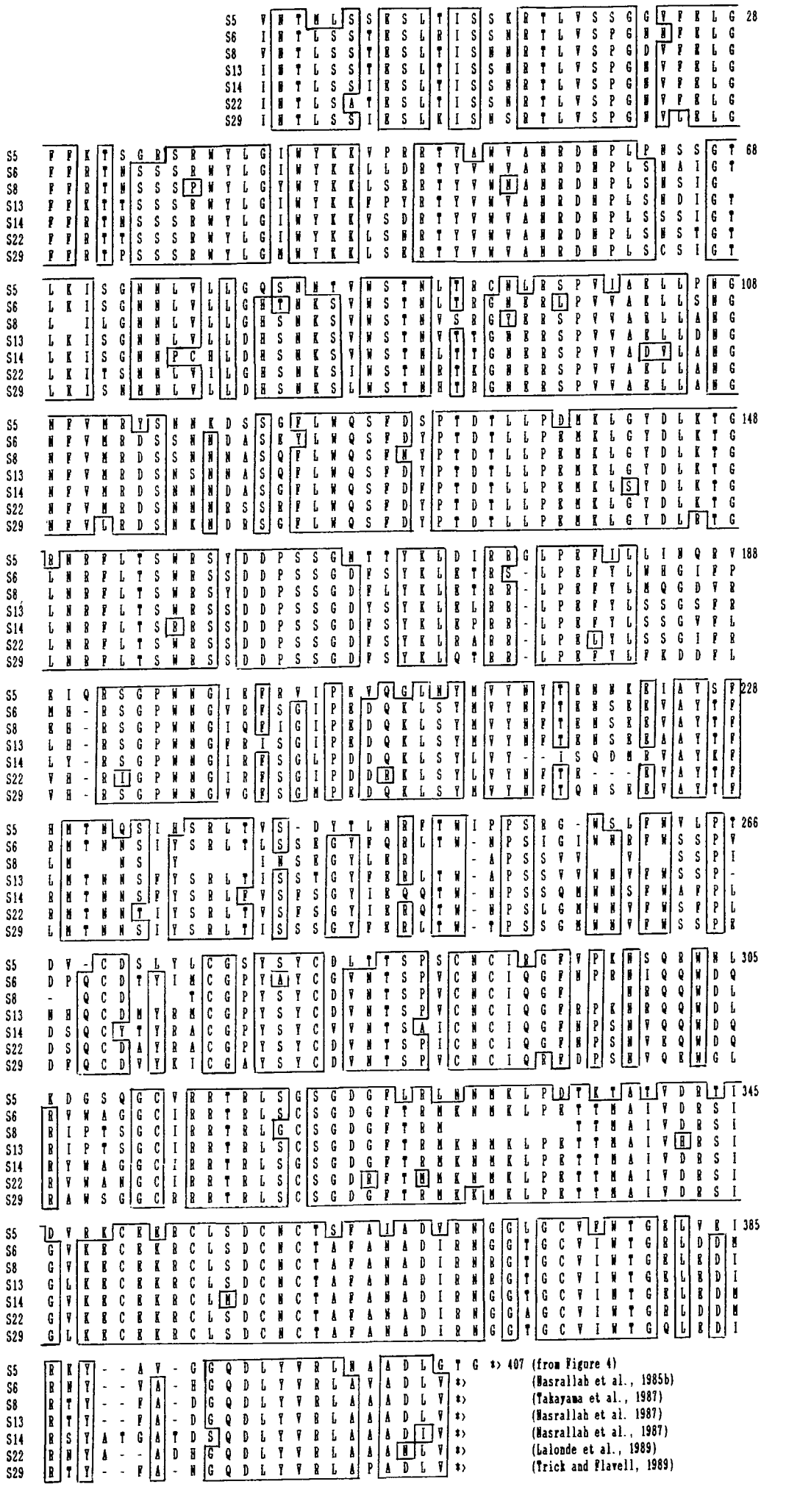
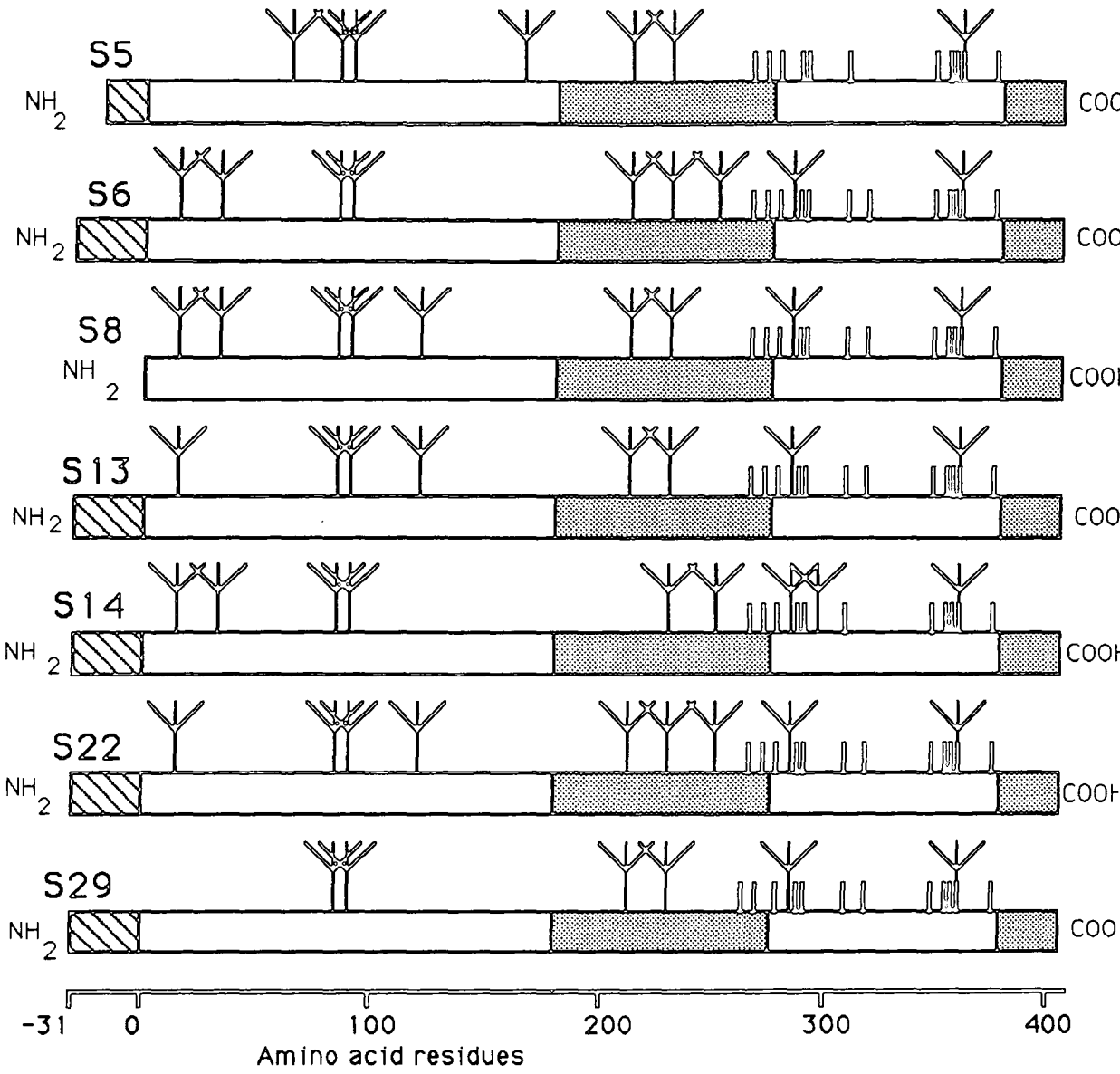
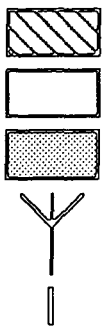

S5
S6
S8
S13
S14
S22
S29
(from Figure 4)
(Hasrallab et al., 1985b)
(Takayama et al., 1987)
(Hasrallab et al. 1987)
(Hasrallab et al., 1987)
(Lalonde et al., 1989)
(Trick and Flavell, 1989)

Figure 6

A Comparison of Known Putative SLSG Sequences From *Brassica*



References: S5- Fig. 4; S6- Nasrallah *et al.* (1985b); S8- Takayama *et al.* (1987); S13 and S14- Nasrallah *et al.* (1987); S22- Lalonde *et al.* (1989); S29- Trick and Flavell (1989).



Putative signal peptide (incomplete for the S5 sequence)

Region of high mean sequence homology (ca. 80%)

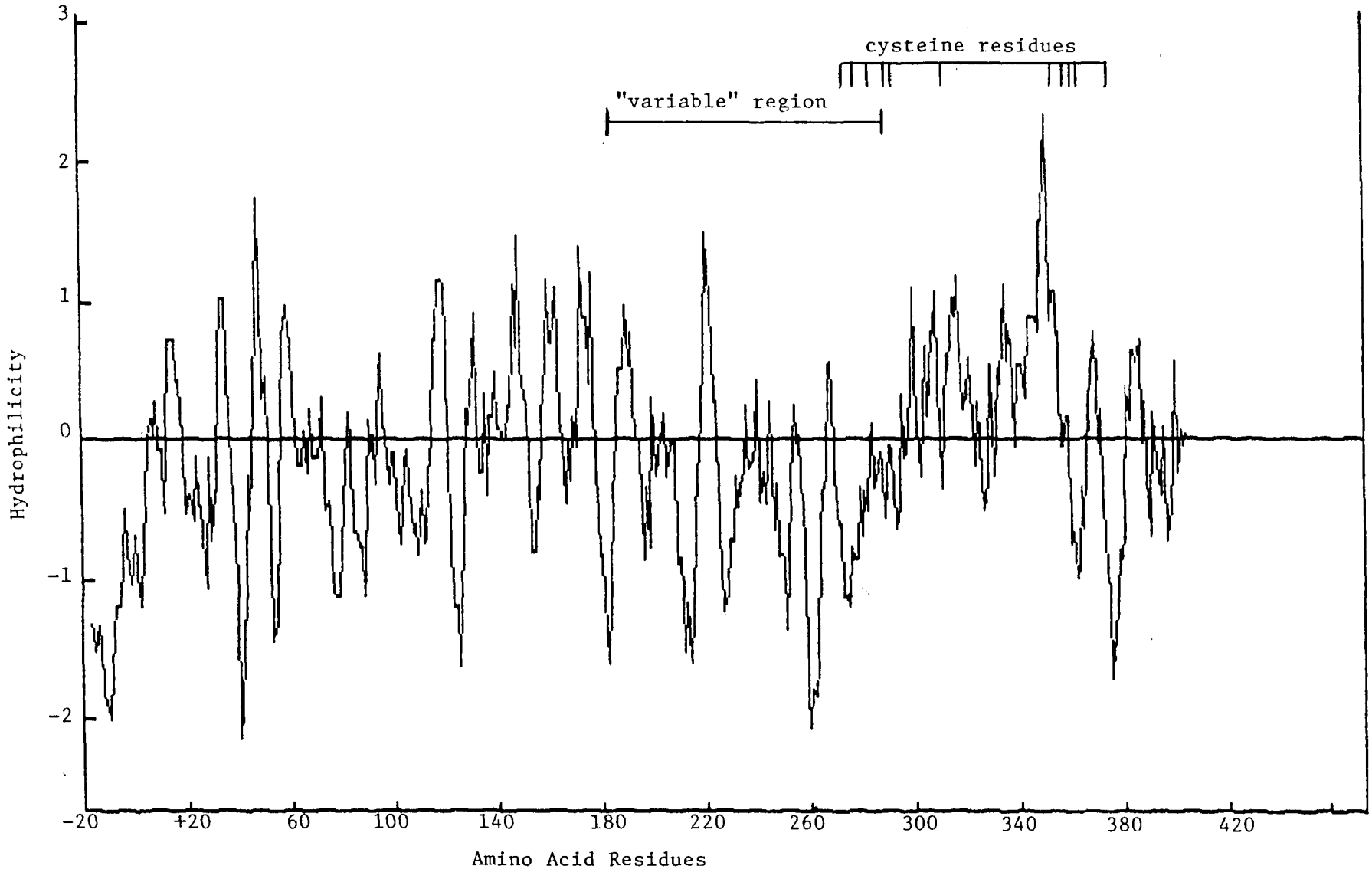
Region of low mean sequence homology (ca. 45%)

Putative N-glycosylation site

Conserved cysteine residue

FIGURE 7

Hydrophilicity plot of the predicted protein from the pBS2 cDNA. The predicted amino acid residues are numbered from the start of the mature protein. The bar labelled "variable region" refers to residues 182 to 275 of the predicted mature protein which show 40% homology to the equivalent region of the S_6 SLSG sequence (Nasrallah et al., 1985b). The values were calculated using the parameters of Kyte and Doolittle (1982) with a span setting of seven residues.


cysteine residues
"variable" region
3
2
1
0
-1
-2
Hydrophilicity
-20
+20
60
100
140
180
220
260
300
340
380
420
Amino Acid Residues

folded glycoprotein and speculate that it may, therefore, be involved in intermolecular recognition events in SI.

The cysteine residues in the protein encoded by the pBS2 cDNA occur mainly in two clusters (from residue 268 to 290 and from residue 349 to 375). The former of these is a predominantly hydrophobic region (Figure 7) whilst the latter is amphipathic. The high level of conservation of the cysteine residues suggests they are important components in determining a basic structure to the SLSG molecule which is common to all its forms. Some or all of these cysteine residues are presumably involved in disulphide bond formation. Note, however, that in the S_5 and S_{14} forms an odd number of conserved cysteine residues is present and so either they are not all involved in disulphide bond formation, or such bonding occurs between putative SLSG molecules (to form covalently bound dimers or polymers).

4.2.2 The Genomic Organization of pBS2 cDNA-Homologous Sequences

The organization of genomic sequences in *Brassica* related to the pBS2 cDNA sequence has been investigated by Southern blotting (Figures 8 and 9). These blots were probed with an EcoRI-SalI fragment of the cDNA which does not contain the anomalous 5' end of the sequence which is possibly derived from a different cDNA. Figure 8 shows an autoradiograph of a blot of DNA from *B. oleracea* lines of two varieties (var. *gemmifera* and var. *alboglabra*), both of which possess the S_5 SI phenotype. RFLPs between the two are clearly identifiable using the pBS2 cDNA probe following complete restriction with EcoRI and HindIII, though not with

FIGURE 8

Autoradiographs of a genomic Southern hybridization of the pBS2 cDNA to restriction fragments of *Brassica oleracea* genomic DNA. Blot A was exposed after low stringency washing (2 x SSC, 65°C) and blot B was exposed after high stringency washing (0.1 x SSC, 65°C). Tracks 1, 2 and 3 contain *B. oleracea* var. *gemmifera* S_5 DNA; tracks 4, 5 and 6 contain *B. oleracea* var. *alboglabra* S_5 DNA. Tracks 1 and 4, EcoRI restrictions; tracks 2 and 5, HindIII restrictions; tracks 3 and 6 KpnI restrictions. Tracks i to iv contain 1, 2, 3 and 4 gene copy equivalents, respectively, calculated from a genome size estimation for *B. oleracea* of 1c = 0.9 pg (Bennet and Smith, 1976). Southern transfer was performed following electrophoresis of the samples on a 0.7% agarose gel.

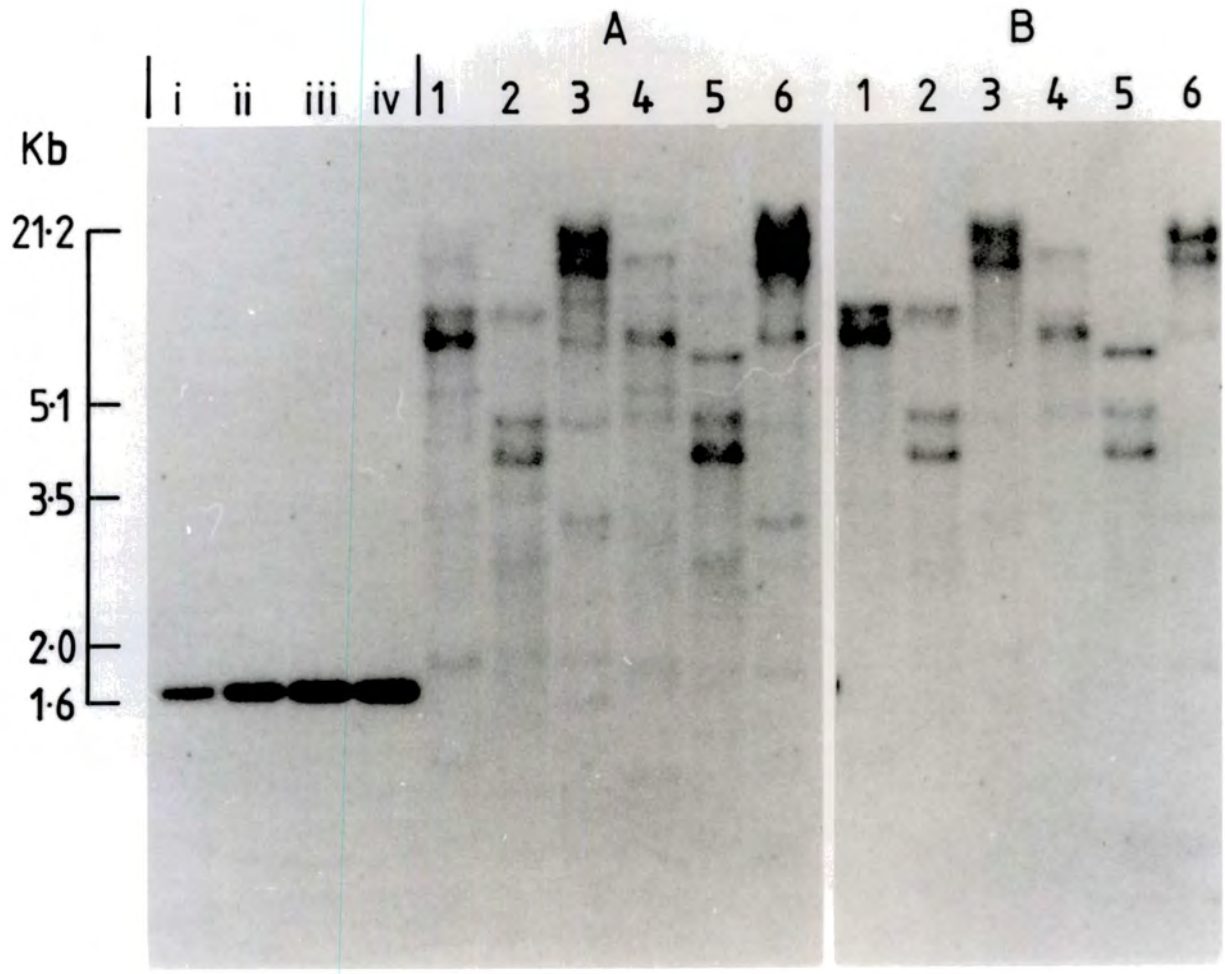
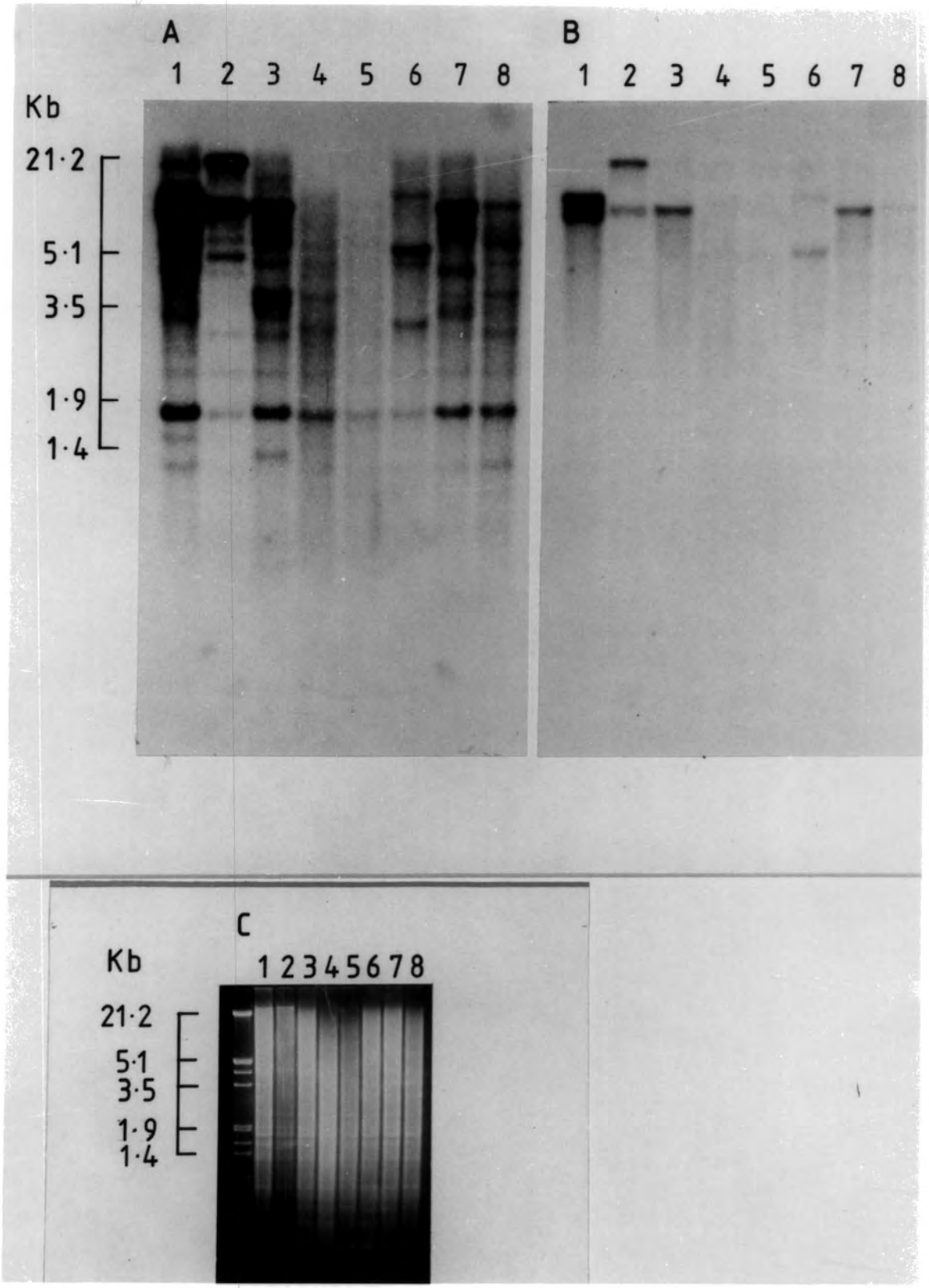

A
B
i ii iii iv 1 2 3 4 5 6 1 2 3 4 5 6
Kb
21·2
5·1
3·5
2·0
1·6

FIGURE 9

Autoradiographs of a genomic Southern hybridization of the pBS2 cDNA probe to genomic DNA from eight self-incompatible lines of *Brassica oleracea*. The genomic DNA samples (5μg per track) were restricted with EcoRI. The hybridization was exposed: A, after low stringency washing (2 x SSC, 65°C); and B, after high stringency washing (0.1 x SSC, 65°C). Genomic DNA samples from the following *B. oleracea* lines were used: track 1, var. *gemmifera* S_5; track 2, var. *alboglabra* S_5; track 3, var. *acephala* S_9; track 4, var. *acephala* x *italica* S_{16}; track 5, var. *capitata* x *acephala* S_{18}; track 6, var. *acephala* S_{20}; track 7, var. *gemmifera* S_{39}; track 8, var. *italica* S_{58}. The samples were electophoresed on a 0.7% agarose gel prior to Southern transfer; the gel is shown (C), stained in ethidium bromide.


A
1 2 3 4 5 6 7 8
Kb
21·2
5·1
3·5
1·9
1·4
B
1 2 3 4 5 6 7 8
C
Kb
1 2 3 4 5 6 7 8
21·2
5·1
3·5
1·9
1·4

KpnI. A number of sequences homologous to the probe are revealed, though many of these are eliminated by high stringency washing (Figure 8B). The copy-number equivalents included indicate only one to two copies of genes with homology to pBS2 to be present on any hybridising fragment of genomic DNA. Figure 9 shows an autoradiograph of a Southern blot of genomic DNA from various lines of *B. oleracea* of known SI genotype. This shows a number of RFLPs generated by EcoRI restriction between the various inbred lines. The lesser homology to the probe of many of the hybridising bands is demonstrated by high stringency washing which reveals only two bands in the case of the S_5 sample and no more than one band in the remaining tracks. This evidence, taken together with the genomic reconstruction incorporated in Figure 8 suggests only one or two copies of the transcribed gene encoding the pBS2 cDNA to be present in the *B. oleracea* genome.

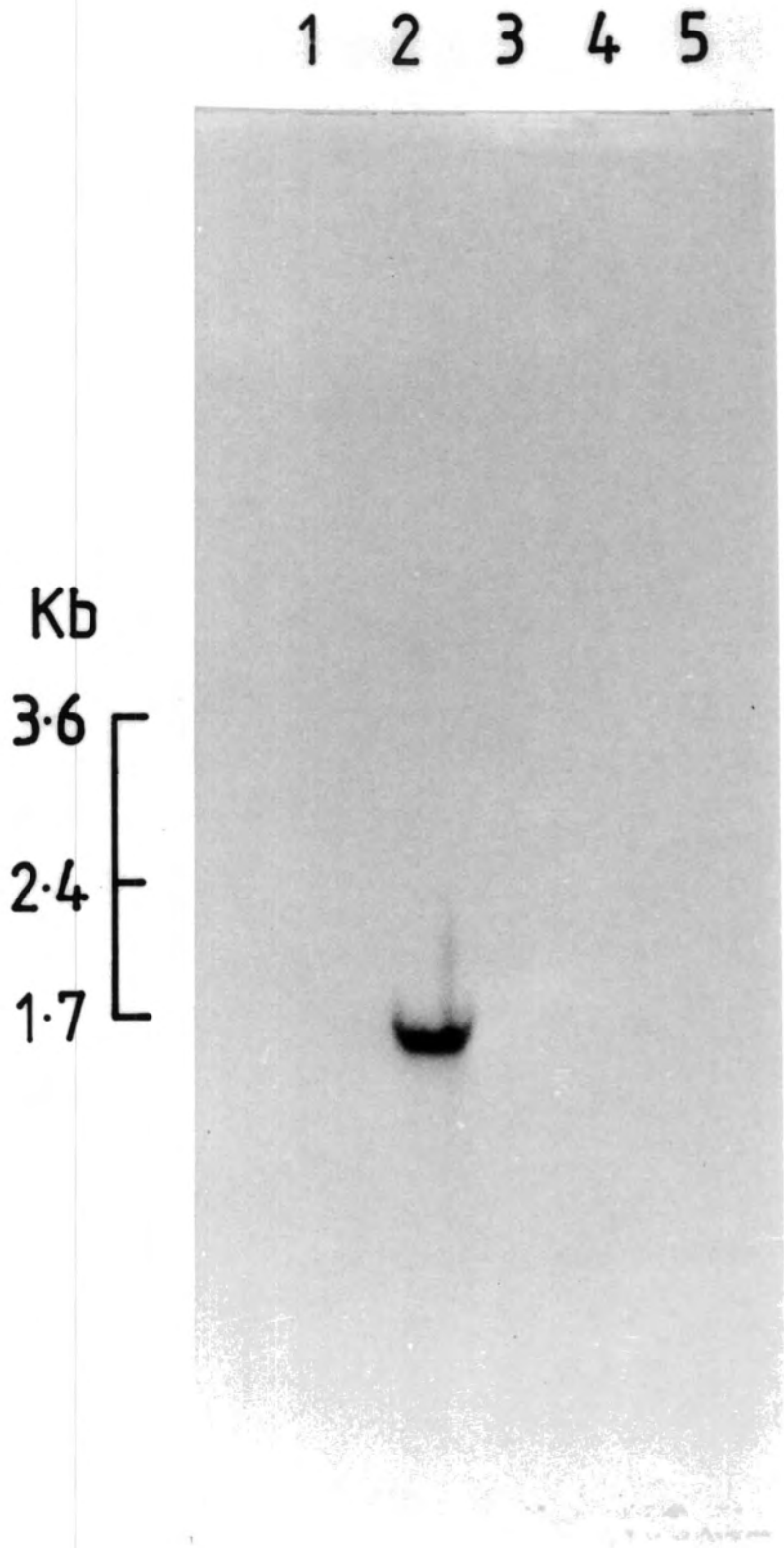
The low level of hybridization of the pBS2 cDNA to S_{16} and S_{18} genomic DNA samples (Figure 9, tracks 4 and 5) is probably artifactual. A limited amount of degradation is shown in the relevant tracks in Figure 9C which reflects difficulties in the DNA extraction procedure caused by the rather old leaf tissue from which these particular samples were obtained.

4.2.3 Expression of Sequences Homologous to the cDNA from pBS2

Northern hybridization (Figure 10) demonstrates the pBS2 cDNA to represent a stigma-specific mRNA of approximately 1600 b. The mRNA was not detectable in leaf

FIGURE 10

Autoradiograph of a Northern hybridization of the pBS2 cDNA to total RNA extracts from plants of an inbred line of *Brassica oleracea* var. *alboglabra* S_5. RNA samples (8μg per track) are from the following tissues: track 1, leaf; track 2, stigma (one day prior to anthesis); tracks 3 to 5, anthers from flower buds of length: 1-2mm, 2-3mm and 3-5mm, respectively. The hybridization was washed at low stringency (2 x SSC, 42^{o}C).


1 2 3 4 5
Kb
3·6
2·4
1·7

tissue or in anther tissue between stages of development correlating with flower bud lengths of between one and five millimeteres. Fluorescence microscopy indicates tetrads to be present within anthers of 1 mm long buds and the tapetum to have degraded at the developmental stage correlating with a bud-length of 5 mm (Figure 11). This suggests the timing of male S-gene expression to be within the studied developmental period as the male S-determinant is regarded as being a tapetal product (Heslop-Harrison, 1967). Operation of the pollen SI system in the material used in the preparation of the Northern blot (Figure 10) was confirmed by cross- and self-pollinations of the plants from which tissues were harvested for RNA extraction (Figure 12).

The pBS2 cDNA probe (a putative S-gene-encoding sequence) is capable of detecting no homologous message, even at low stringency, in anther tissue over the developmental period examined. This is taken as tentative evidence that different, though closely linked, genes encode the SI specificities of pollen grains and the stigma in *Brassica*. It is, however, acknowledged that the timing of S-gene expression in anthers may be even earlier than the period investigated and that such gene expression may be at a very low level and, therefore, not detectable by Northern blotting.

4.2.4 The Nature of the pBS2 cDNA

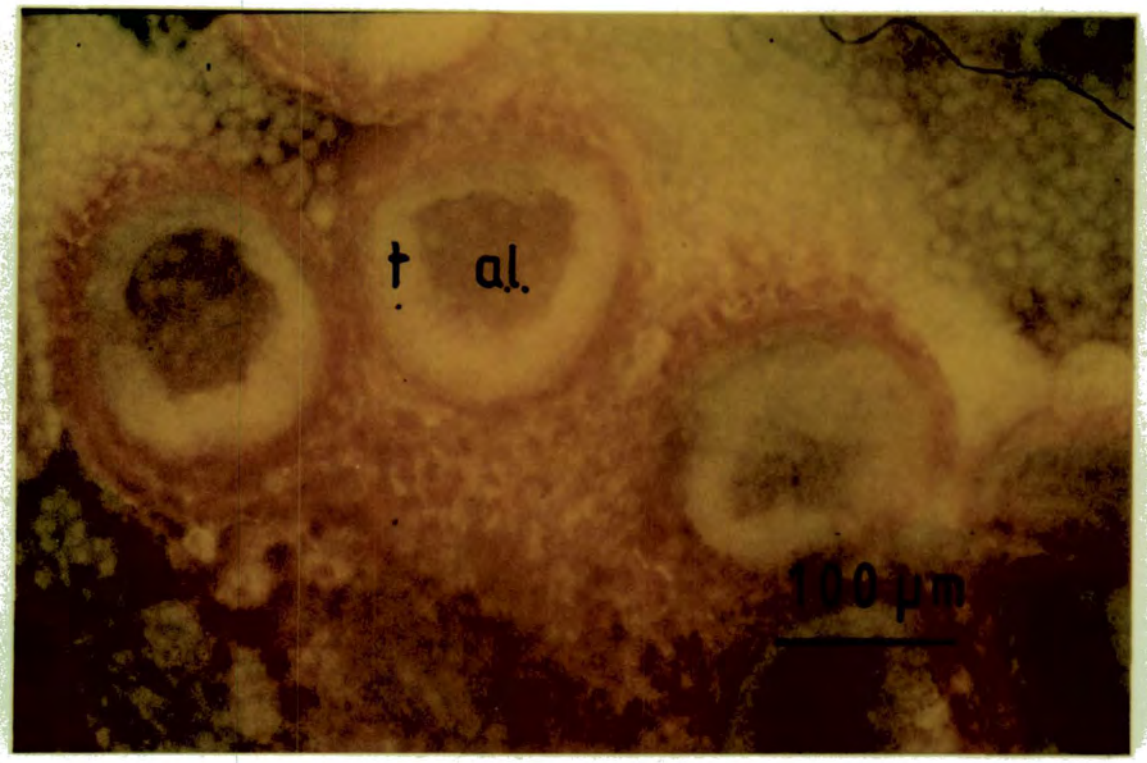
It has been shown (Figure 4) that the pBS2 cDNA from *B. oleracea* S_5 plants is partially homologous to the S_6 SLSG gene (Nasrallah *et al.*, 1988) and appears to be one member of a class of sequences, each derived from a separate line

FIGURE 11

Fluorescence micrographs of transverse sections of anthers of an inbred line of *Brassica oleracea* var.*gemmifera* S_5. The sections were from developmental stages correlating with whole bud lengths of: A, 1mm; and B, 6mm. The tissues were stained in ANS and photographed using fluorescence optics on a Nikon Diaphot microscope fitted with a blue excitation filter. Over the developmental stages between those illustrated (A and B), bud lengths of *B. oleracea* var. *gemmifera* and *B. oleracea* var. *alboglabra* plants were found to be closely similar.

t = tapetum; a.l. = anther loculus.

A



B

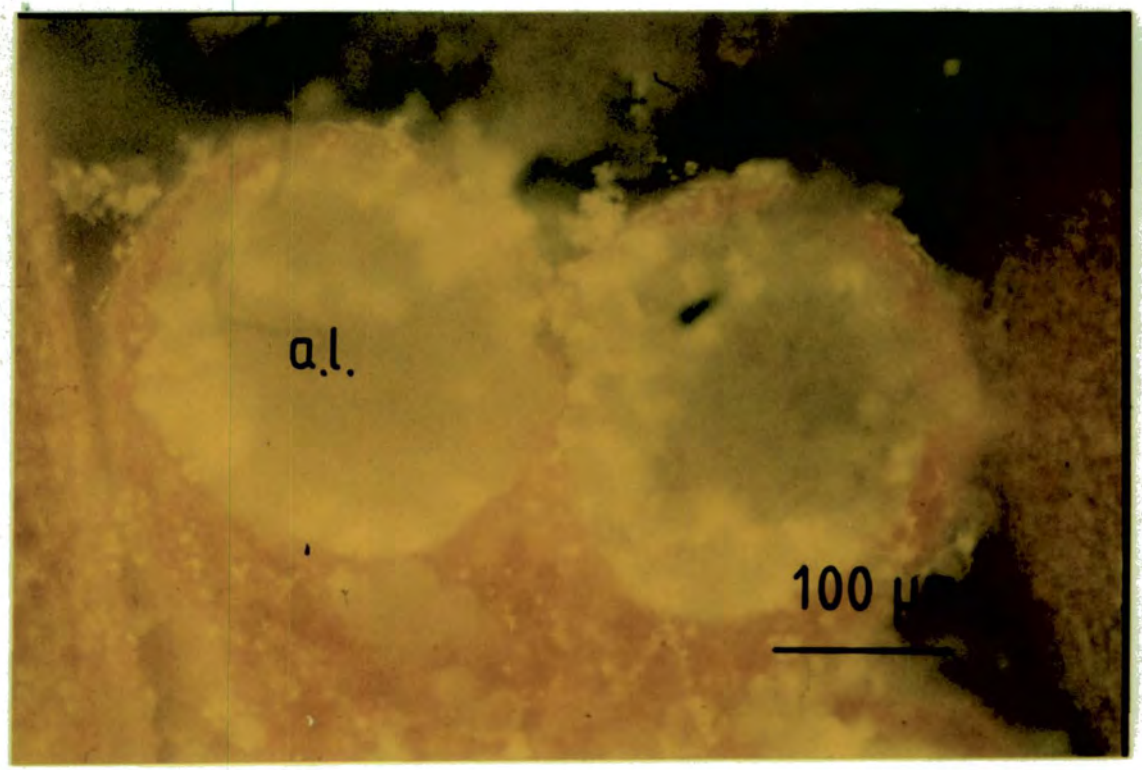
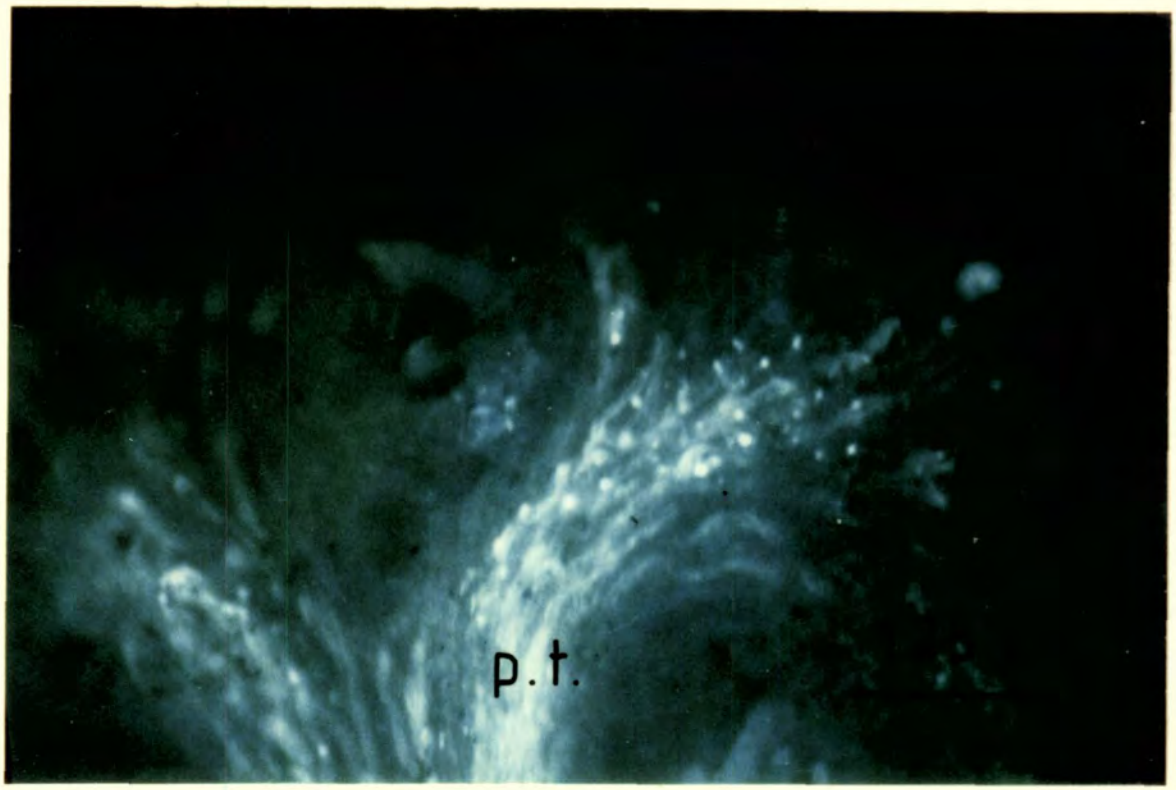


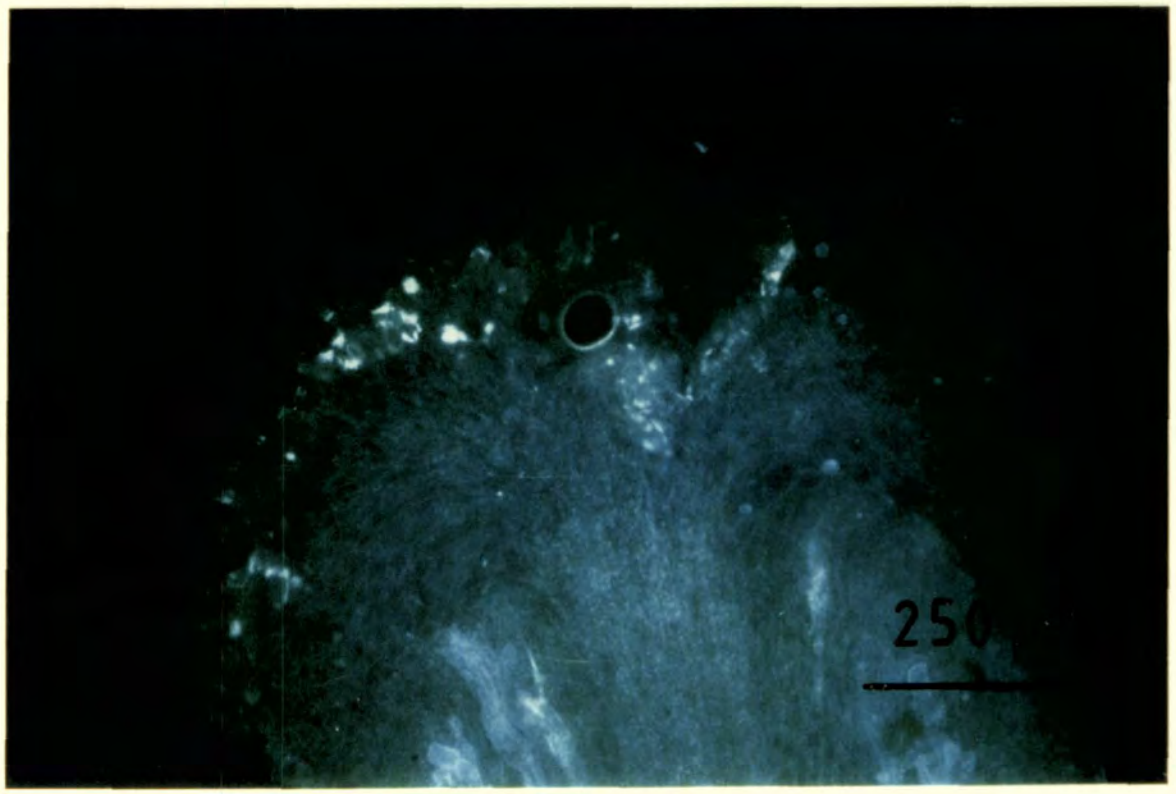
FIGURE 12

Fluorescence micrographs of radial longditudinal sections of the stigma and upper style of plants of an inbred line of *Brassica oleracea* var. *alboglabra* S_5. The plants were pollinated with: A, pollen from an inbred line of *B. oleracea* var. *alboglabra* S_{29}; and B, self pollen. Pollinations were performed 24 h prior to micrography. The tissues were stained in decolourised aniline blue and photographed using fluorescence optics on a Nikon Diaphot microscope fitted with a blue excitation filter.
p.t. = pollen tubes; s.p. = stigmatic papillae.

A



B



of *Brassica* plants with a distinct SI genotype. These sequences show a number of strikingly similar structural features (Figure 6) and have similar tissue and temporal regulation of expression. On the basis of these lines of evidence it is argued that the pBS2 cDNA sequence may represent a form of the SLSG gene. However, this gene has been defined only in the case of the S_6 line of *B. oleracea* and all of the other SLSG-like sequences (including pBS2) can only be regarded as putatively being allelic to this sequence. Particular caution is required in the case of the pBS2 sequence as it shows significantly less homology to the S_6 SLSG gene than do the remainder of the SLSG-like sequences. No RFLP analysis has been published showing any of the putative SLSG sequences (S_5, S_{13}, S_{14}, S_{22} or S_{29}) to be S-linked.

4.2.5 The Basis of S-Allele Specificity in *Brassica*

Two major features of the SLSG molecules have been the focus of speculation regarding their different specificities in (purportedly) bringing about the SI reaction. These are their relatively variable central regions and different positions of putative N-glycosylation signals (Nasrallah *et al.*, 1977, Ebert et al., 1989). The evidence concerning the role of glycosylation in SI has already been reviewed (Sarkar *et al.*, 1989) in Chapter one. This and the known role of glycosyl moieties in other receptor-binding processes suggests the glycosylation of the SLSG molecule to be important in determining the specificity of its interaction.

Several factors may affect the array of glycosyl

moieties displayed by the SLSG molecules: the position of potential glycosylation sites on the SLSG amino acid sequence; the particular glycosylation signals which are selected by the enzymes of glycosylation; the types of glycosyl moieties attached; and the folding of the glycoprotein (which will affect the final positions and orientations of the glycosyl moieties).

The positions of potential glycosylation sites along the amino acid chain of the purported SLSG molecules is the only one of the above factors known for the seven sequences available. The array of potential glycosylation sites is different in each case and this, therefore, may be an important factor in determining S-specificity.

Little is known of the enzymes responsible for glycosylation of the SLSGs. It is known, however, that two different types of oligosaccharides are attached to the S_8 molecule and that not all potential sites are occupied (Takayama *et al.*, 1987). A theoretical consideration becomes relevant here: if the enzymes of glycosylation were responsible for the control of S-specificity by the selective glycosylation of the SLSG molecule, then the SLSG gene would not need to be S-linked, nor would it need to show S-allele-specific sequence variation. Instead, the glycosylation enzymes would need to be encoded by S-linked genes and over fifty distinct combinations of these would be required in order to generate the necessary number of different S-specificities known to exist in *Brassica*. The fact that the SLSG gene is S-linked and shows substantial sequence divergence between lines of different S-specificity suggests control of S-specificity to reside with the SLSG

and not with enzymes responsible for its co-translational, or post-translational processing.

The variability in sequence of the purported SLSG alleles presumably gives rise to molecules with some, though possibly slight, structural diversity. Differences in molecular structure between SLSG molecules may affect the orientations of attached glycosyl moieties radiating from the protein core, even in the case of oligosaccharides bound to the protein in equivalent positions along the sequence in the molecules under consideration. Oligosaccharides linked to residues in the central, more variable region of the SLSG may be particularly affected by this. Thus, the hypothesis that overall sequence divergence plays a role in determining S-specificity may be incorporated within a model which proposes that it is the configuration of glycosyl moieties displayed by the SLSG molecule which is directly involved in the SI recognition reaction.

4.2.6 The Generation of Novel S-Specificities

The SI system in *Brassica* operates with over fifty specificities. It is hypothesized that these are encoded by the SLSG gene (at least in stigmas) and so new S-specificities must result from changes in the SLSG sequence. Two general methods of sequence change might be envisaged: recombination and point mutation. Scrutiny of the available purported SLSG sequences (Figure 5) does not provide any evidence that recombination events have been involved in the generation of allelic differences. If the variable (unboxed) regions are considered, it can be seen that homology exists between two or more of the sequences at

many positions. However, no extensive regions of homology are observed to 'exchange' between pairs of sequences such that, for example, the S_5 and S_6 forms might be homologous over a region in which they are both markedly different from the S_8 SLSG, whilst at some other domain, the S_6 and S_8 molecules share a common sequence, but might be markedly different from the S_5 form. The observed differences in sequence can be accounted for by point mutation assuming that many substitutions may, necessarily, be conservative such that the same mutation may have occurred and been preserved through molecular evolution separately in the generation of more than one S-allele.

If only one S-gene is involved in the SI recognition reaction and its product is expressed in both male and female tissues in *Brassica*, then a new allelic specificity would result from a novel form of the S-gene product caused by mutation at the S-locus. This novel S-gene product would, necessarily, retain the ability to dimerise (or polymerise) in the SI recognition reaction, thereby initiating the sequence of events leading to the inhibition of self-pollen development. However, if the evidence presented of the lack of expression of the SLSG gene in anther tissue is accepted, (Figure 10) (and if it is accepted that the SLSG gene is an S-gene), then a more complex model of the generation of new allelic specificities must be constructed, in order to allow for the control of SI by at least two separate genes. If separate S-genes operate in male and female tissues, then a novel S-specificity would result from a change in one of these followed by a complementary change in the other.

By way of analogy, it may be helpful to consider the specific recognition systems which have arisen in host-pathogen interactions. In these the recognition components are encoded by two separate genomes: one by the plant host and the other by the pathogen, and so there can be no possibility that recognition involves dimerisation of two identical gene products. The plant recognizes the pathogen by some plant-produced component which interacts with a surface borne molecule specific to the pathogen (Albersheim and Anderson-Prouty, 1975). This specific recognition leads to the induction of a hypersensitive reaction (Day, 1974) produced by phytoalexins, resulting in localized necrosis of plant tissues in the region of pathogenic attack, thereby limiting the spread of the pathogen in the host's tissues. Evolution in the parasite would tend to result in a change in its host-pathogen recognition molecule which would prevent recognition and allow pathogenesis to occur. Subsequent evolution in the plant genome may restore specific host-pathogen recognition by a change in the plant's recognition molecule which would complement the change which had occurred in the pathogen's recognition molecule.

If evolutionary rates in plants are high enough to restore specific recognition in host-pathogen interactions, then there is no *a priori* reason why evolutionary changes should not also restore self-recognition in SI systems by a change in one S-gene in order to complement a novel form of the other which may have arisen. Thus, there is no reason to abandon a 'two S-gene' model of SI in *Brassica* on the basis of evolutionary considerations.

4.3 A cDNA Related to an S-Linked Sequence from *Brassica*

4.3.1 Sequence Analysis of cDNA Clone 1E6

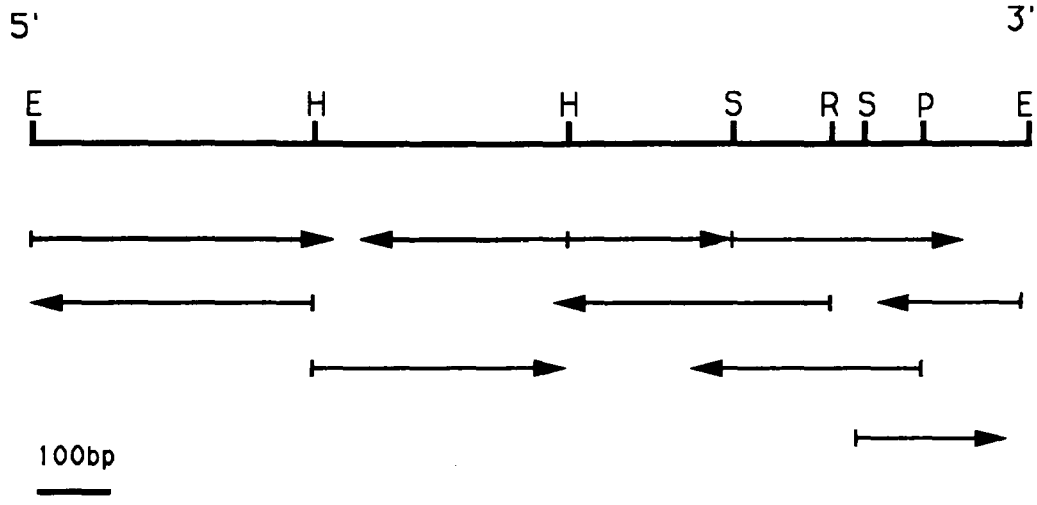
The cDNA from clone 1E6 represents the most abundant species of mRNA found to be tissue-specifically expressed in *Brassica* stigmas. This class of clones was selected by differential screening of a stigma cDNA library with total cDNA probes synthesized from stigma and leaf mRNA preparations (Section 4.1.2). A restriction map of this cDNA was constructed and used as a basis for the sequencing of recombinant M13 clones into which fragments of the 1E6 cDNA had been sub-cloned. The sub-clones generated for sequencing purposes and the extent to which the sequence was read from each of these is shown in Figure 13.

The sequence of the 1E6 cDNA shows it to be a full-length cDNA of 1429 bp, containing an open reading frame of 1332 b (444 amino acids). The coding strand bears a 5' untranslated sequence of 5 b upstream of a putative ATG start codon, a 3' untranslated sequence of 59 b and a poly(A) tail. No putative polyadenylation signal sequence is present in the 3' untranslated region (Figure 14).

The 1E6 cDNA is virtually identical to the sequence published by Lalonde *et al.*, (1989) for a gene termed SLR1 from *Brassica oleracea* plants homozygous for the S_{22} allele. The only difference between these two sequences in their coding regions is the result of a single base substitution (from A to G) which results in arginine rather than lysine at residue 407 of the mature protein encoded by the 1E6 cDNA. The glycoprotein encoded by the SLR1 gene has been independently sequenced (Isogai *et al.*, 1988) with the

Figure 13

Restriction and Sequencing Map of cDNA Clone 1E6

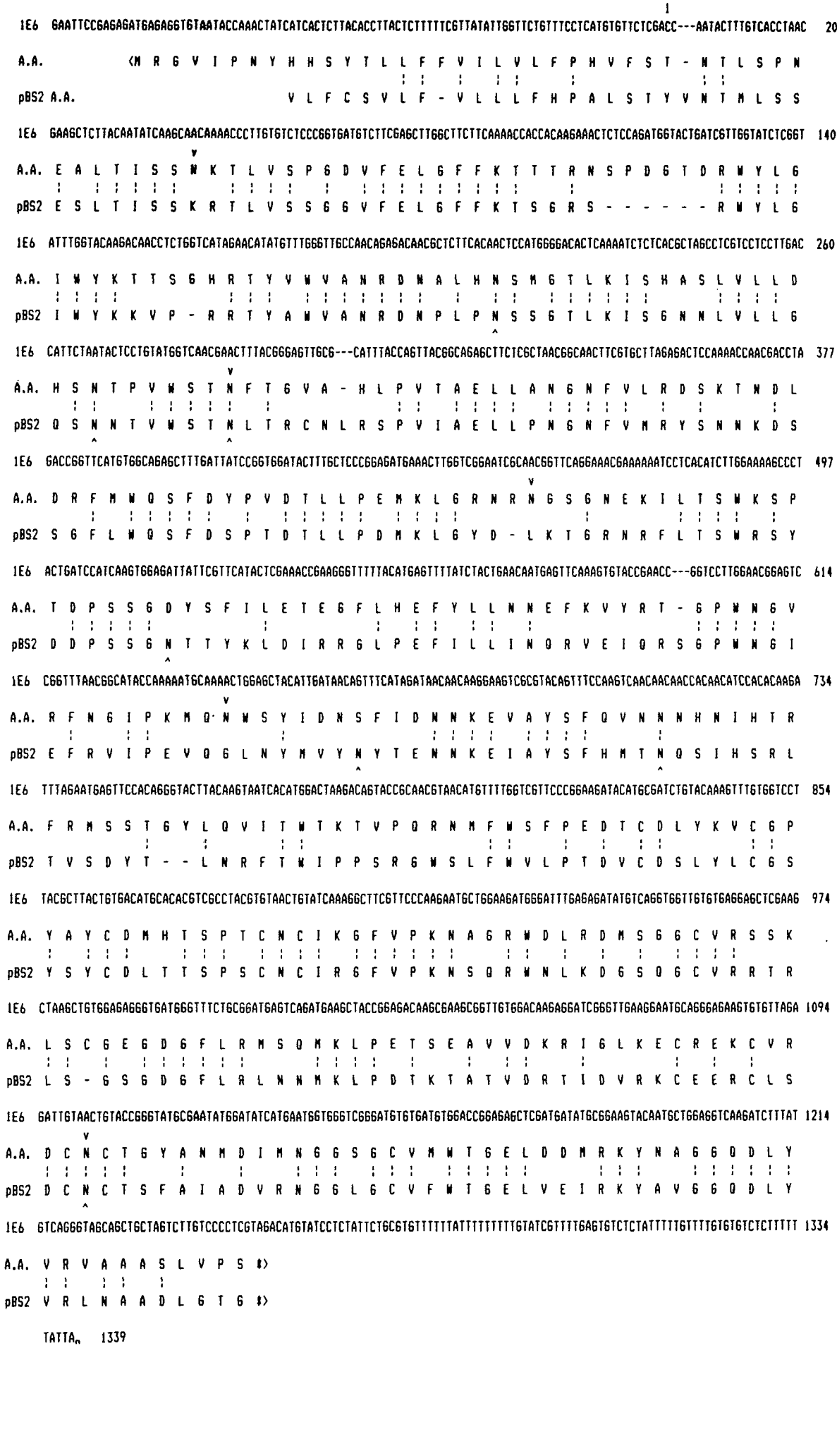


The horizontal bar represents cDNA clone 1E6. Restriction endonuclease sites used in subcloning for sequence analysis are indicated by letters above the bar according to the key shown below. The arrows represent nucleotide sequence data derived from individual recombinant M13 subclones.

E = Eco RI
P = Pvu II
S = Sac I
R = Eco RV
H = Hinc II

FIGURE 14

The nucleotide sequence of cDNA 1E6. The predicted amino acid sequence is shown below and compared with that of the pBS2 cDNA. Putative N-glycosylation sites are shown by arrowheads. The position of the first codon of the mature protein is numbered 1.


1E6 GAATTCCGAGAGATGAGAGGTGTAATACCAAACTATCATCACTCTTACACCTTACTCTTTTTCGTTATATTGGTTCTGTTTCCTCATGTGTTCTCGACC---AATACTTTGTCACCTAAC 20
A.A. (M R G V I P N Y H H S Y T L L F F V I L V L F P H V F S T - N T L S P N
pBS2 A.A. V L F C S V L F - V L L L F H P A L S T Y V N T M L S S
1E6 GAAGCTCTTACAATATCAAGCAACAAAACCCTTGTGTCTCCCGGTGATGTCTTCGAGCTTGGCTTCTTCAAAACCACCACAAGAAACTCTCCAGATGGTACTGATCGTTGGTATCTCGGT 140
A.A. E A L T I S S W K T L V S P G D V F E L G F F K T T T R N S P D G T D R W Y L G
pBS2 E S L T I S S K R T L V S S G G V F E L G F F K T S G R S - - - - - - R W Y L G
1E6 ATTTGGTACAAGACAACCTCTGGTCATAGAACATATGTTTGGGTTGCCAACAGAGACAACGCTCTTCACAACTCCATGGGGACACTCAAAATCTCTCACGCTAGCCTCGTCCTCCTTGAC 260
A.A. I W Y K T T S G H R T Y V W V A N R D W A L H N S M G T L K I S H A S L V L L D
pBS2 I W Y K K V P - R R T Y A W V A N R D W P L P N S S G T L K I S G W W L V L L G
1E6 CATTCTAATACTCCTGTATGGTCAACGAACTTTACGGGAGTTGCG---CATTTACCAGTTACGGCAGAGCTTCTCGCTAACGGCAACTTCGTGCTTAGAGACTCCAAAACCAACGACCTA 377
A.A. H S N T P V W S T N F T G V A - H L P V T A E L L A N G N F V L R D S K T W D L
pBS2 Q S N N T V W S T N L T R C N L R S P V I A E L L P N G W F V W R Y S N W K D S
1E6 GACCGGTTCATGTGGCAGAGCTTTGATTATCCGGTGGATACTTTGCTCCCGGAGATGAAACTTGGTCGGAATCGCAACGGTTCAGGAAACGAAAAAATCCTCACATCTTGGAAAAGCCCT 497
A.A. D R F W W Q S F D Y P V D T L L P E W K L G R N R N G S G N E K I L T S W K S P
pBS2 S G F L W Q S F D S P T D T L L P D W K L G Y D - L K T G R N R F L T S W R S Y
1E6 ACTGATCCATCAAGTGGAGATTATTCGTTCATACTCGAAACCGAAGGGTTTTTACATGAGTTTTATCTACTGAACAATGAGTTCAAAGTGTACCGAACC---GGTCCTTGGAACGGAGTC 614
A.A. T D P S S G D Y S F I L E T E G F L H E F Y L L N N E F K V Y R T - G P W N G V
pBS2 D D P S S G N T T Y K L D I R R G L P E F I L L I N Q R V E I Q R S G P W N G I
1E6 CGGTTTAACGGCATACCAAAAATGCAAAACTGGAGCTACATTGATAACAGTTTCATAGATAACAACAAGGAAGTCGCGTACAGTTTCCAAGTCAACAACAACCACAACATCCACACAAGA 734
A.A. R F N G I P K W Q N W S Y I D N S F I D N N K E V A Y S F Q V N N N H N I H T R
pBS2 E F R V I P E V Q G L N Y W V Y N Y T E N N K E I A Y S F H W T N Q S I H S R L
1E6 TTTAGAATGAGTTCCACAGGGTACTTACAAGTAATCACATGGACTAAGACAGTACCGCAACGTAACATGTTTTGGTCGTTCCCGGAAGATACATGCGATCTGTACAAAGTTTGTGGTCCT 854
A.A. F R W S S T G Y L Q V I T W T K T V P Q R N W F W S F P E D T C D L Y K V C G P
pBS2 T V S D Y T - - L N R F T W I P P S R G W S L F W V L P T D V C D S L Y L C G S
1E6 TACGCTTACTGTGACATGCACACGTCGCCTACGTGTAACTGTATCAAAGGCTTCGTTCCCAAGAATGCTGGAAGATGGGATTTGAGAGATATGTCAGGTGGTTGTGTGAGGAGCTCGAAG 974
A.A. Y A Y C D W H T S P T C N C I K G F V P K N A G R W D L R D W S G G C V R S S K
pBS2 Y S Y C D L T T S P S C N C I R G F V P K N S Q R W N L K D G S Q G C V R R T R
1E6 CTAAGCTGTGGAGAGGGTGATGGGTTTCTGCGGATGAGTCAGATGAAGCTACCGGAGACAAGCGAAGCGGTTGTGGACAAGAGGATCGGGTTGAAGGAATGCAGGGAGAAGTGTGTTAGA 1094
A.A. L S C G E G D G F L R W S Q W K L P E T S E A V V D K R I G L K E C R E K C V R
pBS2 L S - G S G D G F L R L N N W K L P D T K T A T V D R T I D V R K C E E R C L S
1E6 GATTGTAACTGTACCGGGTATGCGAATATGGATATCATGAATGGTGGGTCGGGATGTGTGATGTGGACCGGAGAGCTCGATGATATGCGGAAGTACAATGCTGGAGGTCAAGATCTTTAT 1214
A.A. D C N C T G Y A N W D I W N G G S G C V W W T G E L D D W R K Y N A G G Q D L Y
pBS2 D C N C T S F A I A D V R N G G L G C V F W T G E L V E I R K Y A V G G Q D L Y
1E6 GTCAGGGTAGCAGCTGCTAGTCTTGTCCCCTCGTAGACATGTATCCTCTATTCTGCGTGTTTTTTATTTTTTTTTGTATCGTTTTGAGTGTCTCTATTTTTGTTTTGTGTGTCTCTTTTT 1334
A.A. V R V A A A S L V P S *)
pBS2 V R L N A A D L G T G *)
TATTA$_n$ 1339

finding that the mature protein core of the SLR1 glycoprotein does not include the twenty eight N-terminal residues of the amino acid sequence predicted by the 1E6 cDNA. It is proposed that this region acts as a signal peptide. Two further SLR1-like nucleotide sequences have been determined: a cDNA from an S_{29} homozygous line of *B. oleracea* var. *alboglabra* and a genomic sequence from an S_{63} line of the same variety (Trick and Flavell, 1988; Trick and Flavell, 1989). Despite the identity between the coding sequences of the four SLR1-like nucleotide sequences which have been analyzed, there is a remarkable divergence in their 3' non-coding regions. The 1E6 cDNA (Figure 14) and the SLR1 cDNA (Lalonde *et al.*, 1989) are entirely homologous up to ten bases downstream of their stop codons whereupon they diverge radically. The SLR1 cDNA bears a putative polyadenylation signal within this region, whereas the 1E6 cDNA does not.

The nucleotide sequences of the 1E6 cDNA and pBS2 cDNA (a putative SLSG sequence) are 59% homologous and these cDNAs encode glycoproteins of 50% homology. Of particular interest is the conservation of twelve cysteine residues between the predicted protein sequences from the SLR1 cDNA and all putative SLSG cDNAs (except the S_5 and S_{14} forms where eleven cysteines are conserved). The SLR1 cDNA encodes five potential N-glycosylation signals, of which three are conserved between the SLR1 glycoprotein and the S_6, S_{13}, S_{14} and S_{22} forms of the SLSG glycoprotein. It is not, however, known whether these particular sites are glycosylated.

4.3.2 Genomic Organization of 1E6-Homologous Sequences

Figure 15 shows two autoradiographs of Southern blots of *B. oleracea* genomic DNA probed with the 1E6 cDNA probe. Blot A shows a prominent band representing a 10 Kb EcoRI fragment containing sequences homologous to the probe, and a fainter band representing a 1.5 Kb EcoRI fragment. Two HindIII fragments hybridize to the probe forming a doublet of 3.5 Kb. As there are no sites for HindIII within the 1E6 cDNA, this result can be explained in terms of the following possibilities: more than one gene homologous to the probe may be present in the *Brassica* genome; partial methylation of a HindIII site proximal to the SLR gene may have led to incomplete restriction of the genomic DNA; an intron containing a HindIII site may be present within the SLR1 gene.

Blot B (Figure 15) shows the hybridization of the 1E6 cDNA to EcoRI restrictions of genomic DNA from three varieties of *B. oleracea* after washing at (i) low stringency and (ii) high stringency. A number of weakly hybridising bands are observed after low stringency washing, but these are greatly reduced or eliminated at higher stringency. RFLPs are evident between the three plant varieties for 1E6-homologous sequences; most markedly between *B. oleracea* var. *acephala* and the others.

4.3.3 Expression of 1E6-Homologous Sequences

The 1E6 cDNA has been shown by Northern blotting (Figure 16) to represent a gene which is expressed in *Brassica* stigma tissue, but not in leaf, style or young anther tissue. The tissue specificity of expression of this sequence and of the gene represented by the pBS2 cDNA are

FIGURE 15

Autoradiographs of genomic Southern hybridizations of the 1E6 cDNA probe to restriction fragments of genomic DNA from *Brassica oleracea*. Southern transfer was performed following electrophoresis on 0.7% agarose gels.

Blot A shows genomic DNA from *B. oleracea* var. *gemmifera* restricted with: track 1, EcoRI; Track 2, HindIII. The blot was washed at high stringency (0.1 x SSC, 65°C).

Blot B shows genomic DNA restricted with EcoRI from inbred lines of *B. oleracea*: track 1, var. *gemmifera* S_5; track 2, var. *alboglabra* S_5; track 3, var. *acephala* S_9. The blot was exposed: (i), after low stringency washing (2 x SSC, 65°C); (ii), after high stringency washing (0.1 x SSC, 65°C).

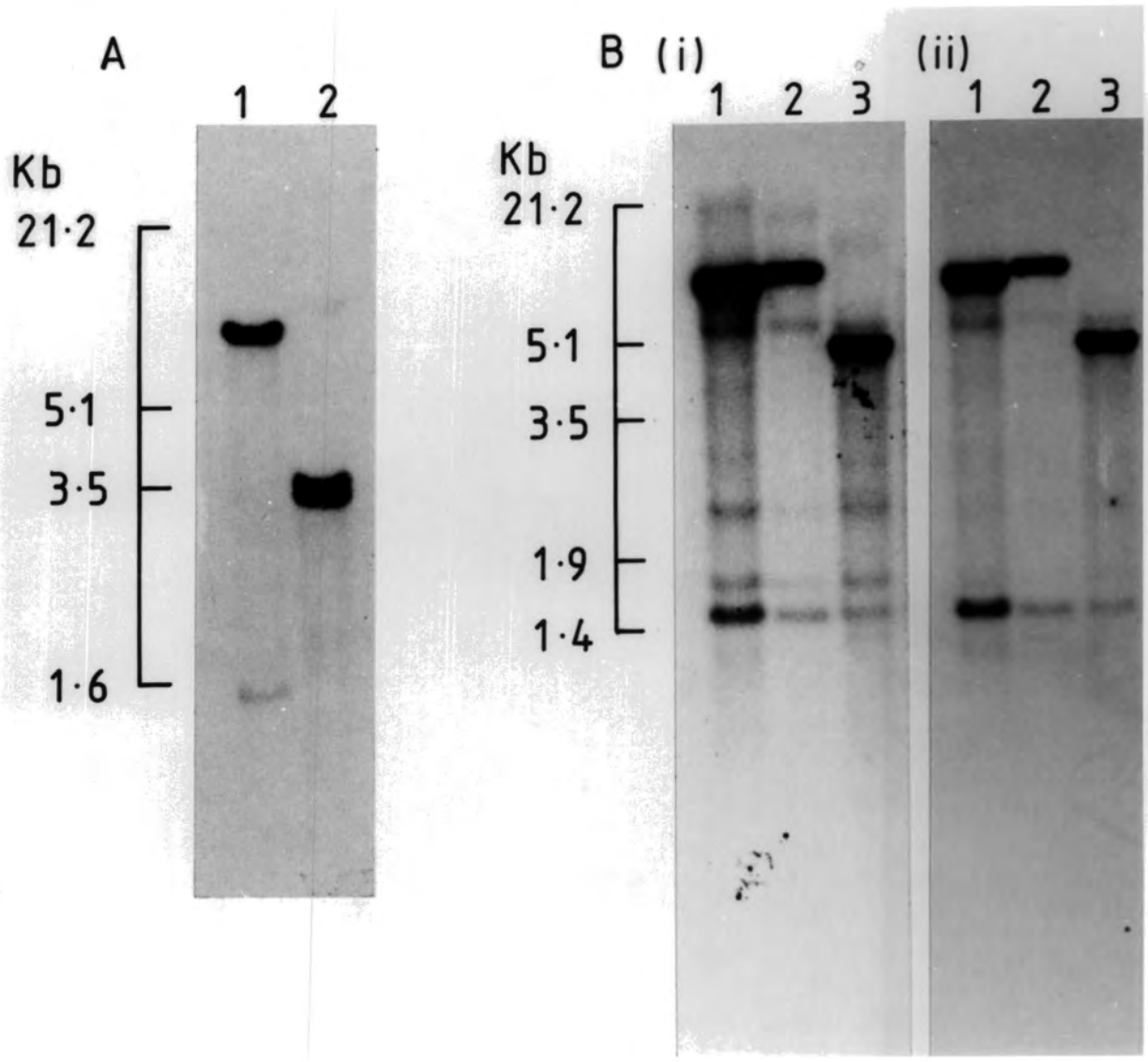
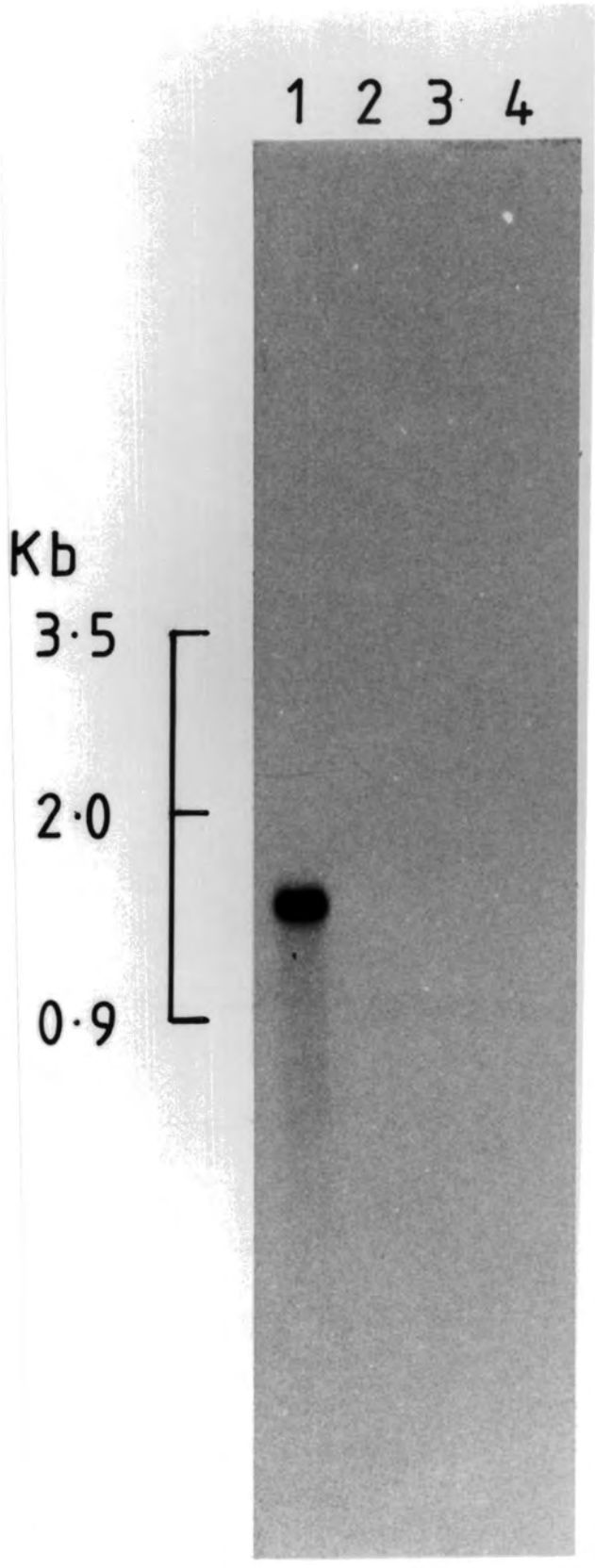

A
1
2
Kb
21·2
5·1
3·5
1·6
B (i)
(ii)
1
2
3
1
2
3
Kb
21·2
5·1
3·5
1·9
1·4

FIGURE 16

Autoradiograph of Northern hybridization of total RNA extracts from plants of an inbred line of *Brassica oleracea* S_5 to the 1E6 cDNA probe. RNA samples (5μg per track) are from the following tissues: track 1, stigma (one day prior to anthesis); track 2, leaf; track 3, style (one day prior to anthesis); track 4, anthers (from flower buds of between 1 and 5mm in length). The blot was washed at high stringency (0.1 x SSC, 60°C).


1 2 3 4
Kb
3·5
2·0
0·9

similar, though from the percentage of clones obtained in cDNA cloning it would appear that message for the former is approximately forty-fold more abundant than message for the latter in stigma tissue one day prior to anthesis.

4.3.4 The Possible Function of the SLR1 Gene

The 1E6 cDNA clearly represents a transcript of the SLR1 gene, the cDNA sequence for which was first published by Lalonde *et al.* (1989). Comparison of cDNA sequences from three different inbred lines of *B. oleracea* demonstrates that the SLR1 gene shows little or no S-specific sequence variability, and Lalonde *et al.* (1989) have shown the SLR1 gene to be not closely linked to the S-locus (by RFLP analysis). The above two lines of evidence (each sufficient in its own) prove that the SLR1 gene cannot be responsible for the specification of SI in *Brassica*. However, the SLR1 and SLSG molecules are similar in structure and expression; their putative signal regions are alike and may both direct the extracellular secretion of the respective glycoproteins. Structural similarities between the two classes of glycoprotein include the identical positioning of cysteine residues and some conservation of putative glycosylation signals. These factors suggest similar types of function for the SLSG and SLR1 glycoproteins.

It has been suggested that the SLSG molecule may be the female SI determinant and as such function as an intercellular recognition component; the SLR1 glycoprotein may function similarly in a parallel recognition process. The SLSG and SLR1 glycoproteins are abundant in stigma tissue and this is suggestive of a structural rather than a

recognition role. However, this may be explained if the further activity of the glycoproteins, following the recognition reactions in which they may participate, requires large quantities of glycoprotein to be present. Another possible explanation is that the glycoproteins may have dual and unrelated roles, one of which requires a large amount of the respective glycoprotein to be present in each case.

4.4 The Evolution of Self-Incompatibility-Like Sequences in *Brassica* and Related Taxa

Figures 17 and 18 show autoradiographs of Southern blots of genomic DNA from species selected as being of various levels of taxonomic divergence from *Brassica oleracea* - the source of the cDNA probes used. Five species of *Brassica* and members of five other genera within the Cruciferae are included.

Reseda lutea was used as it is a self-incompatible member of the Resedaceae - a family placed within the same order as the Cruciferae (Rendle, 1967 and Heywood, 1978). *Senecio viscosus* genomic DNA is included in these experiments as this species has a functionally similar SI system to *Brassica* (i.e. a sporophytic SI system), though it is a member of the Compositae and therefore would be regarded as a distant relative of *Brassica*.

All of the DNA samples included in the hybridizations presented in Figures 17 and 18, with the exception of *Arabidopsis*, are from genera know to contain self-incompatible members (Fryxell, 1957). SI in the

FIGURE 17

Autoradiographs of a genomic Southern hybridization of the pBS2 cDNA to genomic DNA samples of several plant species. The DNA samples (5μg per track) were restricted with EcoRI. Genomic DNA samples were taken from plants of the following: track 1, *Brassica oleracea* var. *gemmifera* S_5; track 2, *B. carinata*; track 3, B. *juncea*; track 4, B. *nigra*; track 5, B. *rapa*; track 6, *Arabidopsis thaliana*; track 7, *Crambe maritima*; track 8, *Hesperis matrionalis*; track 9, *Iberis amara*; track 10, *Raphanus sativus* c.v. Long White Icicle; track 11, *Reseda lutea*; track 12, *Senecio viscosus*. The hybridization was washed at: (A), low stringency (2 x SSC, 65°C); (B), high stringency (0.1 x SSC, 65°C). The samples were electrophoresed on a 0.7% agarose gel prior to Southern transfer which is shown (C) stained in ethidium bromide.

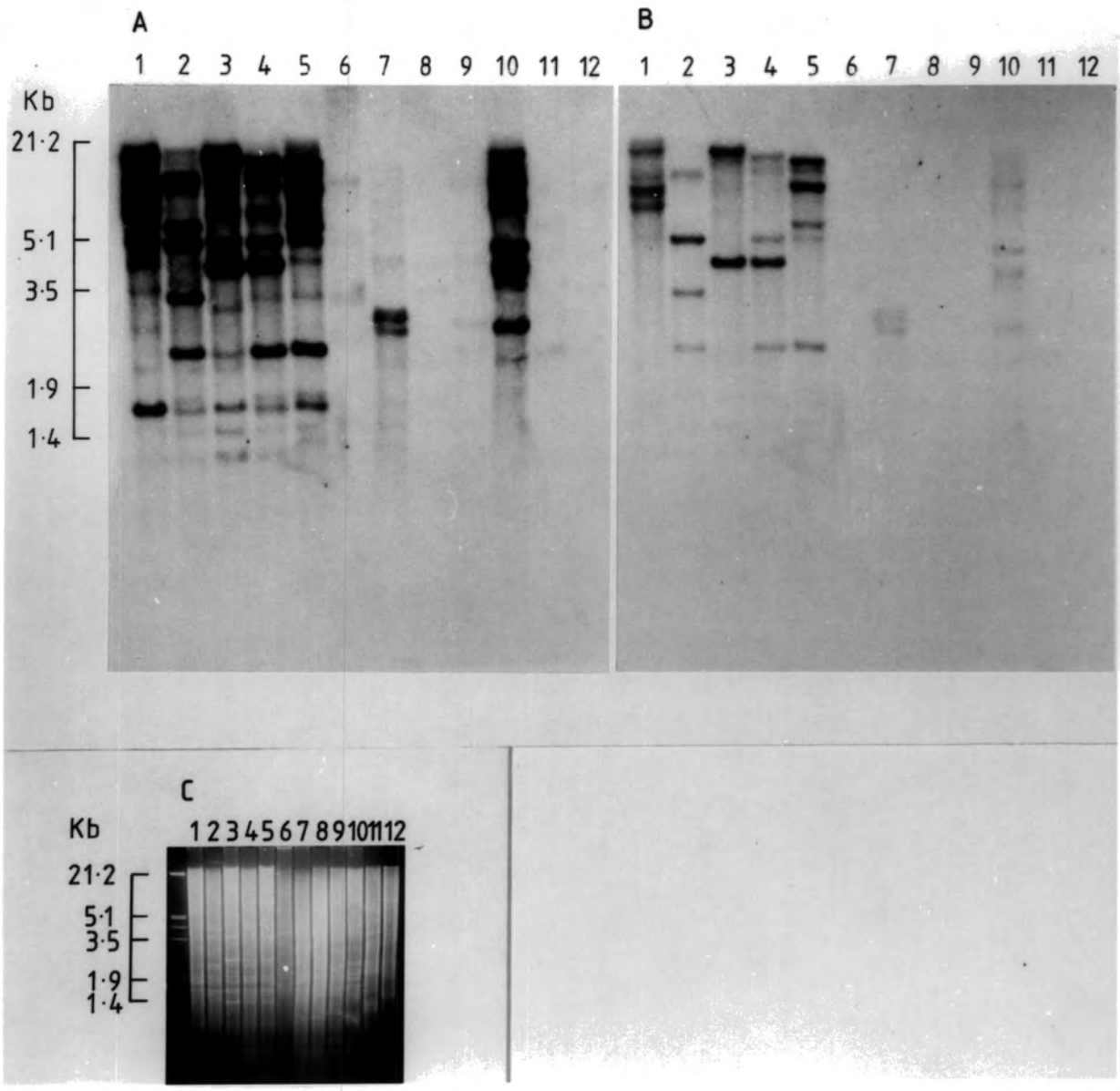
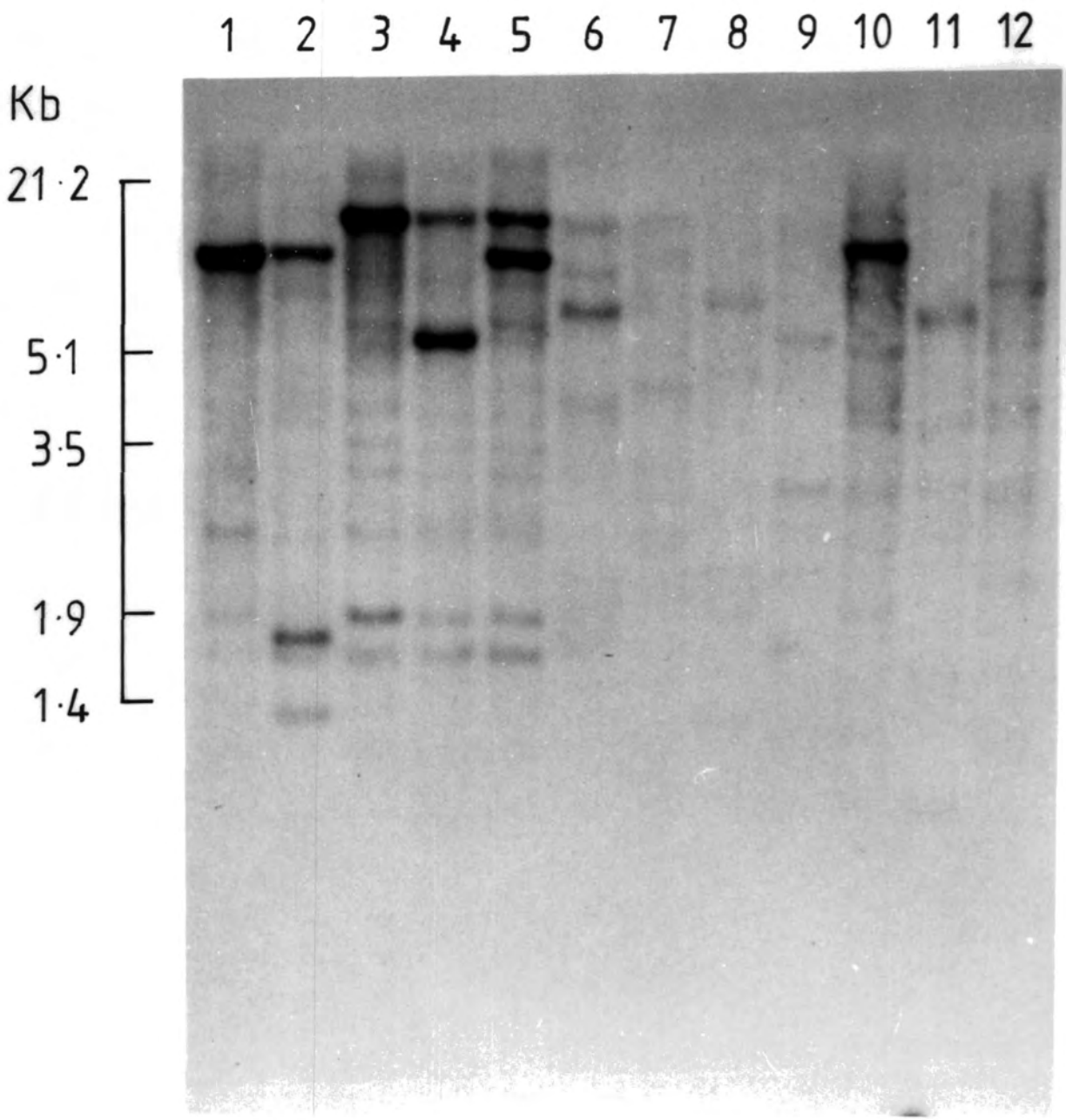

A
1 2 3 4 5 6 7 8 9 10 11 12
Kb
21·2
5·1
3·5
1·9
1·4
B
1 2 3 4 5 6 7 8 9 10 11 12
C
Kb
1 2 3 4 5 6 7 8 9 10 11 12
21·2
5·1
3·5
1·9
1·4

FIGURE 18

Autoradiograph of a genomic Southern hybridization of the 1E6 cDNA to genomic DNA samples from various plant species. The DNA samples (5μg per track) were restricted with EcoRI. The genomic DNA samples were taken from plants of the following: track 1, *Brassica oleracea* var.*gemmifera* S_5; track 2, *B. carinata*; track 3, *B. juncea*; track 4, *B. nigra*; track 5, B. *rapa*; track 6, *Arabidopsis thaliana*; track 7, *Crambe maritima*; track 8, *Hesperis matrionalis*; track 9, *Iberis amara*; track 10, *Raphanus sativus* c.v. Long White Icicle; track 11, *Reseda lutea*; track 12, *Senecio viscosus*. The hybridization was washed at low stringency (2 x SSC, 65°C). The samples were electrophoresed on a 0.7% agarose gel prior to Southern transfer.


1 2 3 4 5 6 7 8 9 10 11 12
Kb
21·2
5·1
3·5
1·9
1·4

experimental material was only tested in the case of *B. oleracea*.

The hybridizations presented in Figures 17 and 18 involve genomic DNA samples which do not represent equivalent amounts of genome (as genomic size estimations are not available for the species used, except in the cases of *Brassica* and *Arabidopsis*). Also, gene copy number estimations have been omitted as these would not be meaningful, except in the case of *B. oleracea*, as where the cDNA probes used are partially diverged from the genomic sequences to which they hybridize, it is not possible to attribute a low signal to either low homology or low copy number alone.

The arguments which can be made from the data presented relate to the simple presence or absence of bands in particular tracks and also to the relative reduction in intensity of bands, representing genomic DNA fragments from different species, which occurs following high stringency washing. The latter type of observation gives some indication of the relative homologies of various hybridising genomic fragments to the *B. oleracea* cDNA probes.

The genomic hybridization shown in Figure 17 is probed with the pBS2 cDNA - a putative SLSG sequence. The blot is washed initially at low stringency (A) and at this stringency genomic DNA fragments are observed to hybridize to the probe in the case of all *Brassica* species and in the cases of *Raphanus* and *Crambe*. The remaining crucifers: *Arabidopsis*, *Hesperis* and *Iberis*, show little hybridization to the cDNA probe, as do *Reseda* and *Senecio*. After washing at high stringency, certain bands persist in the *Brassica*

tracks, though all band intensities are markedly reduced in the *Raphanus* track. *Crambe* also shows a reduction in band intensity on high stringency washing, though less so than in the case of *Raphanus*. It is interesting to note that the intensities of the bands representing the most homologous fragments in tracks corresponding to all of the species of *Brassica* used are reduced in intensity by high stringency washing in equal proportion. This occurs despite the fact that the probe was obtained from the *B. oleracea* line represented in track one of this genomic blot and so the probe might be expected to hybridize more strongly to this sample if it were an SLSG probe exhibiting S-linked sequence polymorphism. This might possibly be an indication that the pBS2 cDNA represents a third class of stigma-specific S-like glycoprotein gene (neither SLSG nor SLR1). Another putative SLSG sequence - BS29-2 (Trick and Flavell, 1989) shows a lack of S-allele-specific hybridization at high stringency in a similar manner to the pBS2 cDNA, though this sequence is more closely related to the remaining putative SLSG alleles than is pBS2.

The genomic hybridization shown in Figure 18 is probed with the 1E6 cDNA (an SLR1 gene probe). The blot is washed at low stringency to allow for the detection of hybridising sequences from diverged taxonomic groups. The probe is observed to hybridize to genomic fragments in the *Brassica* and *Raphanus* tracks (as with the pBS2 probe), but does not hybridize strongly to *Crambe* genomic DNA fragments (in contrast to the pBS2 probe). The 1E6 cDNA probe produces a low signal with *Arabidopsis* genomic DNA and an even lower signal with DNA from the remaining species of crucifers

(though this difference might be accounted for by different genome size rather than homology or gene copy number considerations). Some weak signals are also visible with the 1E6 cDNA probe in the *Reseda* and *Senecio* tracks.

The hybridization results described from Figures 17 and 18 suggest various degrees of phylogenetic divergence from *Brassica oleracea* for the various species examined. All *Brassica* species are indistinguishable in terms of their hybridization to the cDNA probes (though not in terms of the RFLPs they exhibit). *Raphanus* appears to contain separate sequences of homology to both probes and gives banding patterns similar to those obtained for *Brassica* species with both probes in terms of number and positions of bands. These, however, in the case of pBS2-homologous fragments, are clearly of lower homology to the probe than the equivalent *Brassica* fragments. *Crambe maritima* shows moderately strong hybridization to one of the two probes used, whilst other crucifers examined and species from outside the Cruciferae exhibit lower levels of hybridization.

The biosytematic classification of the Cruciferae differs between various authors. Rendle (1967) groups *Brassica* in a separate tribe to *Raphanus* and *Crambe* (which are placed in separate groups within a single tribe). Heywood (1978), however, regards *Brassica*, *Raphanus* and *Crambe* as members of a single tribe, which excludes *Hesperis*, *Iberis* and *Arabidopsis*. The molecular biological data presented here would appear to support the classification of Heywood as correctly reflecting the

phylogenetic divergence of these various genera from *Brassica*.

Homology between SLSG and SLR1 genes is indicative of a gene duplication event. As both putative SLSG and SLR1 cDNA probes (Figures 17 and 18 respectively) hybridize to mutually exclusive sets of genomic fragments in certain genera of the Cruciferae, it is clear that the gene duplication event which led to these sequences preceded the evolutionary divergence of a number of modern genera. It is also clear, from the washing of Southern hybridizations at differential stringencies (Figure 17), that evolution in the SLSG-like sequence has continued since the divergence of the genera *Brassica* and *Raphanus*, by the lower homology of the *Brassica* pBS2 cDNA probe to *Raphanus* genomic DNA fragments than to those of *Brassica* species.

The *Brassica* SLSG gene has been shown to exist in a number of allelic forms. The sequence variability exhibited by the SLSG gene has been explained in terms of its putative function as an S-gene controlling recognition in SI. It is reasonable to suppose, within this hypothesis, that the diversification of the SLSG gene into a number of allelic forms has been driven by natural selection; that is, the fixation rate of mutant forms of SLSG alleles has been high.

In order to quantitate the divergence of the SLSG gene the proportions of active and silent base pair differences between a pair of putative SLSG alleles has been compared. The nucleotide sequences (omitting codons occasionally to maximize homology between the sequences) have been entered into a computer programme (written by Dr. Vaughan Hilder) which yields corrected average divergence percentages

according to the method of calculation of Perler *et al.* (1980). This method calculates rates at which mutations are fixed which result in differences between two sequences for active and silent positions, respectively. (Active mutations result in a change in the translational meaning of a sequence, whilst silent mutations result in synonymous codons.) The method of calculation corrects for multiple mutations at single nucleotide sites (which is why percentage figures may exceed 100%).

Silent mutations are essentially neutral with respect to the environment and are free to occur until the differences at 'silent' nucleotide sites become saturated. Active changes are subject to natural selection, though the strength and direction of this selection may vary. Conservative amino acid substitutions , resulting from such active changes, may be largely neutral with respect to the environment and so be fixed at a high rate. Deleterious active mutations, however, should be subject to negative natural selection such that they cannot become fixed, whereas positively advantageous mutations may be fixed at a higher rate than the fixation of silent mutations.

Figure 19 shows corrected average percentage divergence figures for active (replacement) and silent positions in one conserved and one variable region of the SLSG gene for the putative SLSG alleles from S_5 and S_{29} lines of *B. oleracea*. The sequences are taken from the cDNAs pBS2 (Figure 4) and BS29-2 (Trick and Flavell, 1989). The divergence figures show the ratio of silent to active position differences between the two cDNAs to be approximately 6:1 in the

FIGURE 19

An analysis of the nucleotide sequence divergence between two SLSG-like sequences. The sequences compared are: the pBS2 cDNA from an S_5 inbred line of *Brassica oleracea* (Figure 4); and the BS29-2 cDNA from an S_{29} inbred line of *B. oleracea* (Trick and Flavell, 1989). Percentage weighted average divergence figures are given for "silent" positions (ie. those which result in synonymous codons) and for "replacement" positions (ie. those which result in a change in the encoded amino acid residue). The method of calculation of percentage weighted average divergence is that given by Perler *et al.* (1980), the calculation being made using a computer programme written by Dr. Vaughan Hilder. Two separate pairs of regions of the SLSG-like cDNAs are compared: a pair of relatively conserved regions (S5C V S29C); and a pair of relatively unconserved regions (S5V V S29V).

S5C V S29C: A comparison between relatively conserved regions of the two cDNAs. The nucleotide sequences are labelled: S5C (pBS2 conserved region); and S29C (BS29-2 conserved region). These regions of the cDNAs represent amino acid residues 321 to 370 of the predicted mature proteins. The predicted amino acid sequences derived from the two cDNAs are compared below the nucleotide sequences.

S5V V S29V: A comparison between relatively unconserved regions of the two cDNAs. The nucleotide sequences are labelled: S5V (pBS2 relatively unconserved region); and S29V (BS29-2 relatively unconserved region). These regions of the cDNAs represent amino acid residues 183 to 270 of the predicted mature proteins. The predicted amino acid sequences derived from the two cDNAs are compared below the nucleotide sequences.

S5C v S29C

```
S5C
AGTGGAGATGGGTTTTTGCGGCTAAATAACATGAAGTTGCCGGATACTAAGACTGCGACGGTGGATCGGACCATTGATGT
GAGAAAATGTGAAGAGAGGTGTCTTAGCGATTGTAACTGTACGTCGTTTGCTATTGCGGATGTTCGAAAC

S29C
AGTGGAGATGGTTTTACCAGGATGAAGAAGATGAAGTTGCCAGAAACTACGATGGCTATTGTCGACCGCAGTATTGGTCT
GAAAGAATGTGAGAAAAGGTGCCTTAGCGATTGTAATTGTACCGCATTTGCAAATGCGGATATCCGGAAT

SGDGFLRLNNMKLPDTKTATVDRTIDVRKCEERCLSDCNCTSFAIADVRN
::::: :   :::: :  : ::: :    :: :::::::::: :: :: ::
SGDGFTRMKKMKLPETTMAIVDRSIGLKECEKRCLSDCNCTAFANADIRN
```

Weighted average divergence for 'silent' positions
= 118.708149 %

Weighted average divergence for 'replacement' positions
= 19.1275608 %

S5V v S29V

```
S5V
CTTATAAATCAACGTGTTGAAATAAGGAGCGGTCCGTGGAATGGAATCGAGTTTAGAGTTATACCGGAGGTGCAAGGATT
AAATTACATGGTTTACAATTATACGGAGAACAATAAAGAGATCGCTTACTCGTTCCATATGACCAACCAAAGCATCCACT
CCAGATTGACAGTCAGTGACTATACACTCAATCGATTCACGTGGCCGCCATCACGGGGTTGGAGCCTGTTCTGGGTTTTA
CCAACGGACGTGTGC

S29V
TTTAAAGACGACTTTCTAGTGCATCGGAGTGGTCCATGGAATGGAGTCGGATTTAGTGGTATGCCAGAGGACCAAAAATT
GAGTTACATGGTGTACAATTTCACACAGAATAGTGAGGAGGTCGCTTATACATTTCTAATGACCAACAACAGCATCTACT
CGAGATTGACAATAAGCTCAGGGTATTTTGAGCGACTGACGTGGACTCCGTCATCAGGGTGGAACGTGTTCTGGTCTTCT
CCAGAGGACTTCTGC

LINQRVEIRSGPWNGIEFRVIPEVQGLNYMVYNYTENNKEIAYSFHMTNQSIHSRLTVSDYTLNRFTWPPSRGWSLFWVL
        ::::::::  :   :: : : ::::: : :  : :: : ::: :: :::: :     : :: :: ::  ::
FKDDFLVHRSGPWNGVGFSGMPEDQKLSYMVYNFTQNSEEVAYTFLMTNNSIYSRLTISSGYFERLTWTPSSGWNVFWSS
PTDVC
: : :
PEDFC
```

Weighted average divergence for 'silent' positions
= 93.2704868 %

Weighted average divergence for 'replacement' positions
= 33.9395758 %

conserved region and 3:1 in the variable region. In both of these regions natural selection would appear to have operated predominantly in a conservative manner as active mutations have been fixed at a lower rate than silent mutations. (It is possible to conceive of a situation where novel forms of proteins might confer a very high selective advantage and so the active mutations which cause them might be fixed at a higher rate than silent mutations.) The variable region of the putative SLSG protein would appear, from the divergence statistics, to have evolved at approximately twice the rate of the conserved region in terms of active mutations fixed per unit of time.

The SLR1 gene is virtually identical in *B.oleracea* varieties *gemmifera* (Figure 14) and *alboglabra* (the BS29-1 sequence in Trick and Flavell, 1989). Clearly, the divergence of these varieties has been too recent for either active or silent nucleotide changes to have become fixed to any marked extent.

4.5 A cDNA Encoding a Glycine-Rich Cell Wall Protein Gene Showing Stigma Specific Expression

4.5.1 Analysis of the cDNA Designated 15H11

Clone 15H11 represents the longer cDNA of two homologous sequences found amongst 2000 stigma cDNA clones. These two clones were selected by differential screening of a stigma cDNA library (Section 4.1.2). Clone 15H11 was sequenced following sub-cloning in both orientations in M13 vectors and by sub-cloning both fragments produced by restriction of the cDNA using NcoI such that the sequence of

each could be read from the (internal) NcoI restriction site. In this way the entire nucleotide sequence was read from both strands of the cDNA (Figure 20).

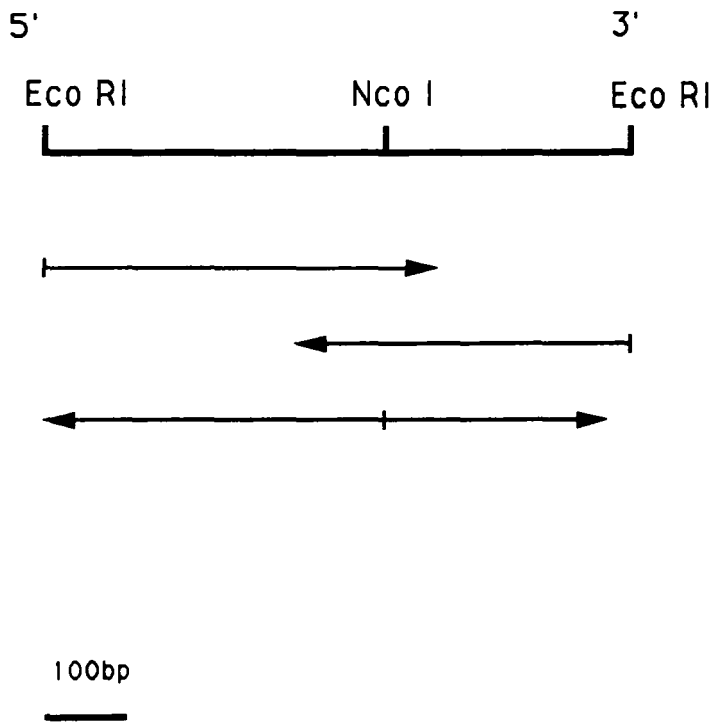
The 15H11 cDNA (Figure 21) is a partial cDNA of 776 bp, the coding strand containing an open reading frame of 660 b (220 amino acids) from its 5' end to an in-frame stop codon. The sequence includes a 3' non-translated sequence of 102 b containing a putative polyadenylation signal, and a poly(A) tail. The predicted protein sequence contains 75% glycine residues, largely as motifs of [Gly.Gly.Gly.X] , where X represents any amino acid. This motif repeats throughout the entire predicted protein sequence.

4.5.2 Genomic Organization of 15H11-Homologous Sequences

The organization of genomic sequences homologous to the 15H11 cDNA has been investigated by Southern blotting (Figure 22A). An autoradiograph of *B. oleracea* genomic DNA restricted with EcoRI and HindIII restriction enzymes respectively is shown, probed with the 15H11 cDNA and washed at two stringencies: (i) in a solution containing 0.1 x SSC at 60^{o}C and (ii) a solution of identical ionic strength but at 65^{o}C. The result indicates that, at the lower stringency, quite a large number of genomic fragments of various sizes hybridize to the probe. However, a considerable reduction in the number of hybridizing fragments is observed on washing at the slightly higher stringency, leaving only two prominent bands in each track. The most prominent bands hybridizing after the blot is washed at low stringency (Figure 22A(i)) are completely

Figure 20

Restriction and Sequencing Map of cDNA Clone 15H11



The horizontal bar represents cDNA 15H11. Restriction endonuclease sites used in subcloning for sequence analysis are shown above this. The arrows represent nucleotide sequence data derived from individual recombinant M13 subclones.

FIGURE 21

The nucleotide sequence of cDNA 15H11. The predicted amino acid sequence derived from the cDNA is shown below. A putative polyadenylation site is indicated by: <polyA+>.

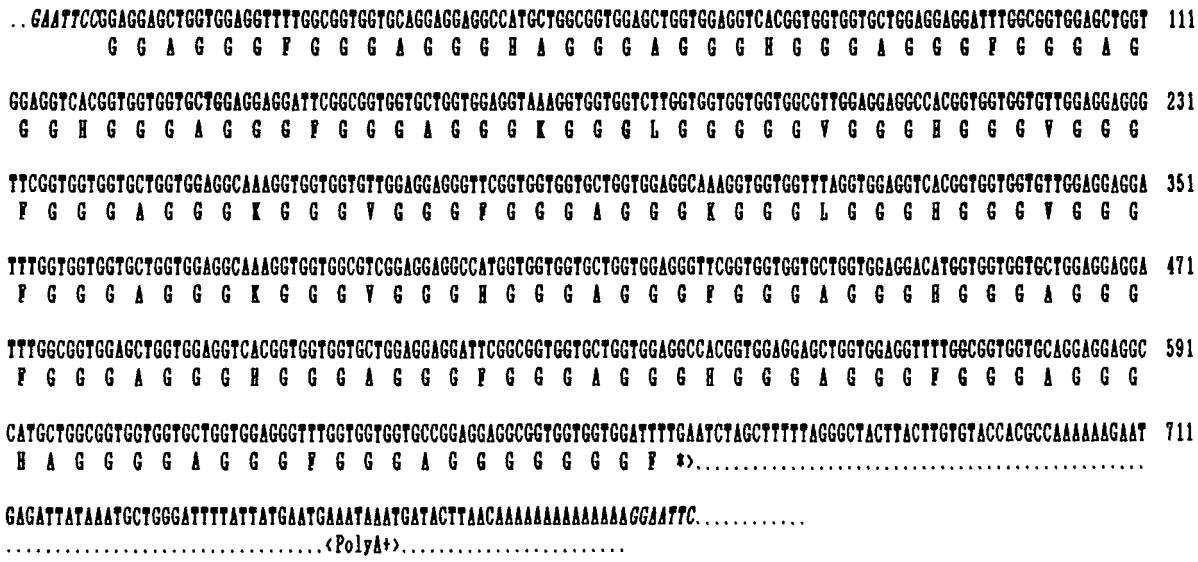

..GAATTCCGGAGGAGCTGGTGGAGGTTTTGGCGGTGGTGCAGGAGGAGGCCATGCTGGCGGTGGAGCTGGTGGAGGTCACGGTGGTGGTGCTGGAGGAGGATTTGGCGGTGGAGCTGGT 111
G G A G G G F G G G A G G G H A G G G A G G G H G G G A G G G F G G G A G
GGAGGTCACGGTGGTGGTGCTGGAGGAGGATTCGGCGGTGGTGCTGGTGGAGGTAAAGGTGGTGGTCTTGGTGGTGGTGGTGGCGTTGGAGGAGGCCACGGTGGTGGTGTTGGAGGAGGG 231
G G H G G G A G G G F G G G A G G G K G G G L G G G G G V G G G H G G G V G G G
TTCGGTGGTGGTGCTGGTGGAGGCAAAGGTGGTGGTGTTGGAGGAGGGTTCGGTGGTGGTGCTGGTGGAGGCAAAGGTGGTGGTTTAGGTGGAGGTCACGGTGGTGGTGTTGGAGGAGGA 351
F G G G A G G G K G G G V G G G F G G G A G G G K G G G L G G G H G G G V G G G
TTTGGTGGTGGTGCTGGTGGAGGCAAAGGTGGTGGCGTCGGAGGAGGCCATGGTGGTGGTGCTGGTGGAGGGTTCGGTGGTGGTGCTGGTGGAGGACATGGTGGTGGTGCTGGAGGAGGA 471
F G G G A G G G K G G G V G G G H G G G A G G G F G G G A G G G H G G G A G G G
TTTGGCGGTGGAGCTGGTGGAGGTCACGGTGGTGGTGCTGGAGGAGGATTCGGCGGTGGTGCTGGTGGAGGCCACGGTGGAGGAGCTGGTGGAGGTTTTGGCGGTGGTGCAGGAGGAGGC 591
F G G G A G G G H G G G A G G G F G G G A G G G H G G G A G G G F G G G A G G G
CATGCTGGCGGTGGTGGTGCTGGTGGAGGGTTTGGTGGTGGTGCCGGAGGAGGCGGTGGTGGTGGATTTTGAATCTAGCTTTTTAGGGCTACTTACTTGTGTACCACGCCAAAAAAGAAT 711
H A G G G A G G G F G G G A G G G G G G G F *>.....
GAGATTATAAAATGCTGGGATTTTATTATGAATGAAATAAATGATACTTAACAAAAAAAAAAAAAAAGGAATTC.....
.....<PolyA+>.....

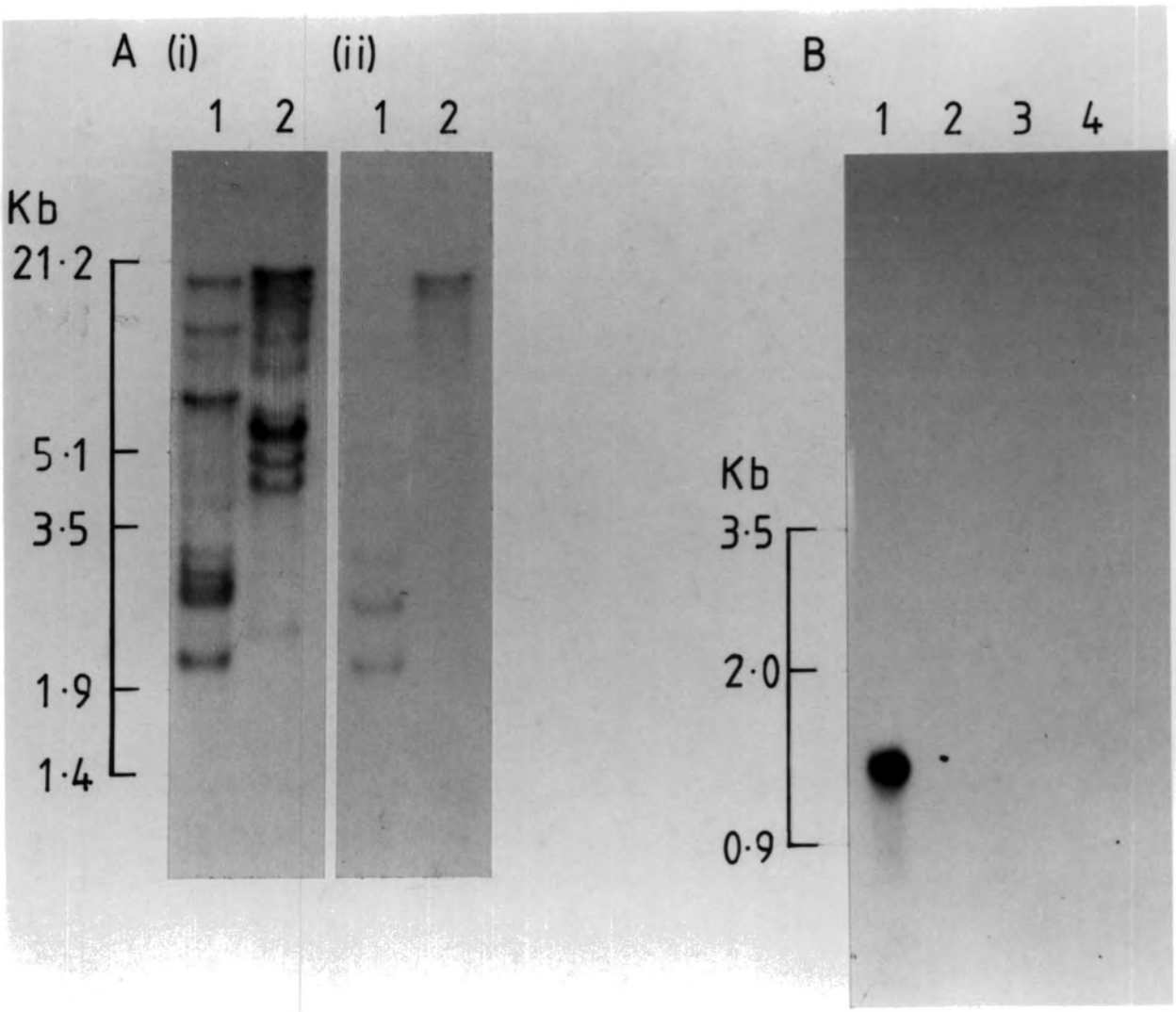
FIGURE 22

A

Autoradiographs of a genomic Southern hybridization of a probe of the 15H11 cDNA to restriction fragments of *Brassica oleracea* var.*gemmifera* S_5 DNA. Genomic DNA samples (5μg) were restricted with: track 1, EcoRI; track 2, HindIII, prior to electrophoresis on a 0.7% agarose gel and Southern transfer. The hybridization was exposed: (i), following washing in 0.1 x SSC at 60°C; and (ii), following washing in 0.1 x SSC at 65°C.

B

Autoradiograph of a Northen hybridization of a probe of the 15H11 cDNA to total RNA samples from plants of *Brassica oleracea* var. *gemmifera* S_5. The RNA samples (5μg per track) were from the following tissues: track 1, stigma (1 day prior to anthesis); track 2, leaf; track 3, style (1 day prior to anthesis); track 4, anthers (from flower buds of between 1 and 5mm in length).


A
(i)
(ii)
1
2
1
2
Kb
21·2
5·1
3·5
1·9
1·4
B
1
2
3
4
Kb
3·5
2·0
0·9

eradicated by higher stringency washing. This indicates that these bands represent a high gene copy number of sequences of less than complete homology to the probe, by comparison to the lower gene copy numbers of the more homologous sequences which continue to hybridize to the probe after high stringency washing.

4.5.3 The Expression of 15H11-Homologous Sequences

The gene encoding cDNA 15H11 shows stigma specific expression by Northern blot analysis (Figure 22B). A ca. 1300 b transcript was detected in stigma tissue but not in leaf, style or young anther tissue.

4.5.4 The Function of the Glycine-Rich Protein Encoded by cDNA 15H11

Glycine-rich protein (GRP) gene expression has recently been demonstrated in a number of plant systems. Reddy and Poovaiah (1987) have measured the accumulation of a GRP in the receptacles of strawberry fruits. They note this to relate, possibly in a casual manner, to the cessation of growth of the receptacle.

Condit and Meagher (1987) have isolated a GRP gene from *Petunia*. They report expression of this sequence in leaves, stems and flowers. The expression of the GRP gene is shown to be enhanced by a factor of approximately twenty five in all tissues by wounding and is also found to be greater in young rather than old tissues. S_1 nuclease mapping indicates a likelihood that the same gene is expressed in all the aforementioned situations.

Two GRP genes with [Gly.Gly.Gly.X] repeating motifs have been sequenced from *Phaseolus vulgaris* (Keller *et al.*, 1988). The expression of one of these has been localized to the vascular tissues of the hypocotyl and ovary and has been found to be enhanced by wounding. Immunological studies demonstrate that antibodies raised to the GRP cross-react with purified cell wall fractions, indicating the GRP to be a cell wall protein. It has been found to be synthesized specifically in protoxylem elements of the vascular system (Keller *et al.*, 1989).

It is interesting to note that GRP genes are related to hydroxyproline-rich cell wall protein genes in that their sequences are partially homologous if the coding strand of one is compared with the anticoding strand of the other. (The GGG codon, encoding glycine, complements the CCC codon, encoding proline.) It has been suggested that one of these two forms of cell wall protein has arisen from the other by gene duplication followed by inversion of the coding sequence (Keller *et al.*, 1988).

The 15H11 cDNA sequence clearly represents a structural protein gene and not an information carrying molecule involved in SI or any similar phenomenon. It has been shown to encode a GRP of similar sequence to those known from *Phaseolus* (Keller *et al.*, 1988) and therefore is almost certainly a cell wall protein-encoding sequence. The gene homologous to the 15H11 cDNA shows stigma specific expression amongst those tissues for which expression was studied. The expression of this sequence in the stigma, but not in the style may relate to the distribution of

protoxylem elements in these tissues, given the evidence of protoxylem-specific expression of GRP genes (Keller *et al.*, 1989), though detailed analysis would be necessary to confirm this speculation.

4.6 A cDNA Representing a Stigma-Specifically Expressed *Brassica* Gene

4.6.1 Sequence Analysis of cDNA 11H4

A cDNA clone designated 11H4 - the longer of the two homologous clones selected by differential screening of a *Brassica* stigma cDNA library of 2000 clones with stigma and leaf total cDNA probes (Section 4.1.2), was sub-cloned into M13 vector DNA for sequence analysis. The nucleotide sequence was read from either end of the cDNA and these sequences were found to overlap, such that the entire nucleotide sequence was determined, though not from both strands of the cDNA (Figure 23).

cDNA 11H4 was found to be a partial cDNA of 465 bp, the coding strand of which encodes an open reading frame of 285 b (95 amino acids) from its 5' end to an in-frame stop codon. The coding strand shows a 3' untranslated region of 179 b containing four putative polyadenylation signals. No poly(A) tail is present. No identification of this cDNA could be made using the available database (Section 3.10).

4.6.2 Genomic Organization of 11H4-Homologous Sequences

Figure 24A shows an autoradiograph of a Southern blot of *Brassica oleracea* genomic DNA probed with the 11H4 cDNA probe. Two prominent bands representing DNA fragments

FIGURE 23

Nucleotide sequence of the 11H4 cDNA. The predicted amino acid sequence derived from the cDNA is shown below. Putative polyadenylation signals are indicated by: <polyA+>.

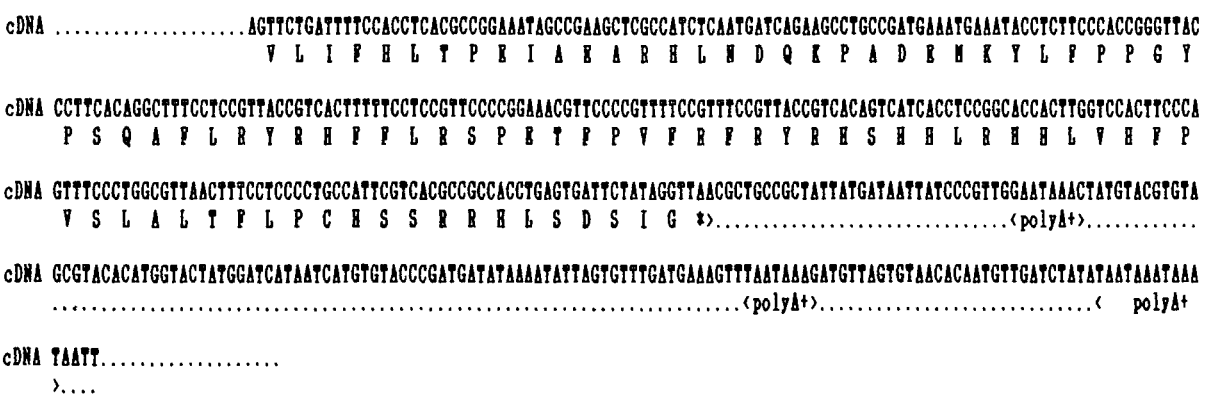

cDNA .....AGTTCTGATTTTCCACCTCACGCCGGAAATAGCCGAAGCTCGCCATCTCAATGATCAGAAGCCTGCCGATGAAATGAAATACCTCTTCCCACCGGGTTAC
V L I F H L T P E I A E A R H L N D Q K P A D E M K Y L F P P G Y
cDNA CCTTCACAGGCTTTCCTCCGTTACCGTCACTTTTTCCTCCGTTCCCCGGAAACGTTCCCCGTTTTCCGTTTCCGTTACCGTCACAGTCATCACCTCCGGCACCACTTGGTCCACTTCCCA
P S Q A F L R Y R H F F L R S P E T F P V F R F R Y R H S H H L R H H L V H F P
cDNA GTTTCCCTGGCGTTAACTTTCCTCCCCTGCCATTCGTCACGCCGCCACCTGAGTGATTCTATAGGTTAACGCTGCCGCTATTATGATAATTATCCCGTTGGAATAAACTATGTACGTGTA
V S L A L T F L P C H S S R R H L S D S I G *>.....<polyA+>.....
cDNA GCGTACACATGGTACTATGGATCATAATCATGTGTACCCGATGATATAAAATATTAGTGTTTGATGAAAGTTTAATAAAGATGTTAGTGTAACACAATGTTGATCTATATAATAAATAAATAAA
.....<polyA+>.....< polyA+
cDNA TAATT.....
>....

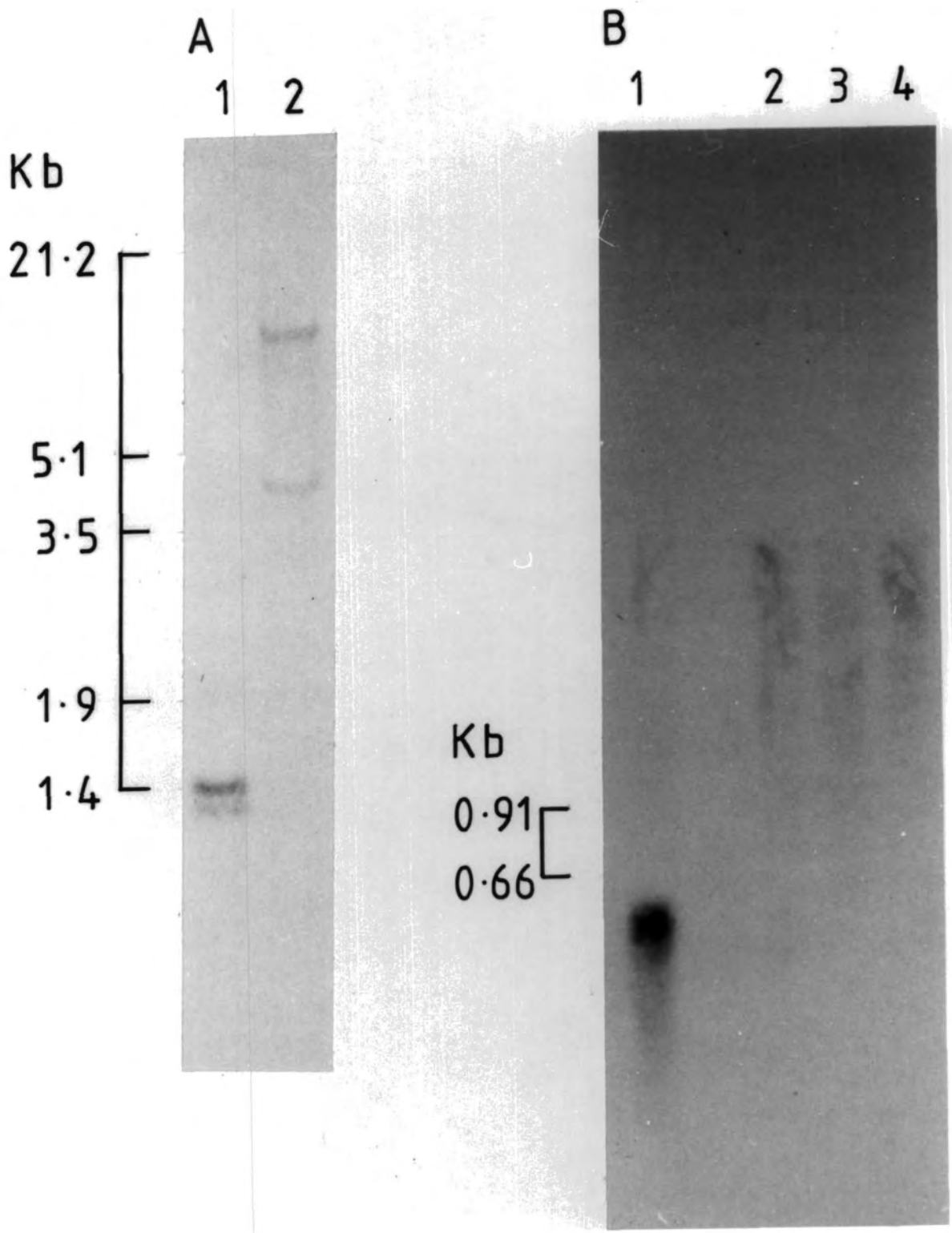
FIGURE 24

A

Autoradiograph of a genomic Southern hybridization of of a probe of the 11H4 cDNA to restriction fragments of *Brassica oleracea* var. *gemmifera* S_5 genomic DNA. The genomic DNA samples (5μg per track) were restricted with: track 1, EcoRI; track 2, HindIII, prior to electrophoresis on a 0.7% agarose gel and Southern transfer. The hybridization was washed at high stringency (0.1 x SSC, 65°C).

B

Autoradiograph of a Northen hybridization of a probe of the 11H4 cDNA to total RNA extracts from plants of *Brassica oleracea* var. *gemmifera* S_5. The RNA samples (5μg per track) were from the following tissues: track 1, stigma (1 day prior to anthesis); track 2, leaf; track 3, style (1 day prior to anthesis); track 4, anthers (from flower buds of between 1 and 5mm in length).


A
1
2
Kb
21·2
5·1
3·5
1·9
1·4
B
1
2
3
4
Kb
0·91
0·66

homologous to the probe are evident in tracks corresponding to DNA restricted with EcoRI and HindIII, respectively.

4.6.3 Expression of 11H4-Homologous Sequences

The expression of the gene encoding cDNA 11H4 has been investigated by Northern hybridization analysis (Figure 24B). A transcript of ca. 520 b was detected in *Brassica* stigma tissue harvested one day prior to anthesis, but not in leaf, style or young anther tissue.

4.7 A cDNA Homologous to mRNA Transcripts Detected in *Brassica* Stigma, Style and Anther Tissue

The cDNA clone designated 13G6(i) was selected from a *Brassica* stigma cDNA library on the basis of strong hybridization to a stigma, but not a leaf, total cDNA probe (Section 4.1.2). The cDNA is approximately 1150 bp in length. No other sequences homologous to cDNA 13G6(i) were detected in the 2000 clone stigma cDNA library. The cDNA was cloned in M13 DNA vectors and partially sequenced. A reading frame could only be assigned, from the nucleotide sequence data obtained, to the 3' end of the cDNA. Figure 25 shows the nucleotide sequence of the 3'end of the clone 13G6(i), together with the C-terminal region of the predicted protein. The coding strand of the cDNA contains a 3' untranslated region of 216 b containing two putative polyadenylation signals. A poly(A) tail is present.

FIGURE 25

Nucleotide sequence of the 13G6(i) cDNA. The predicted amino acid sequence derived from the cDNA is shown below. Putative polyadenylation signals are indicated by: <polyA+>.

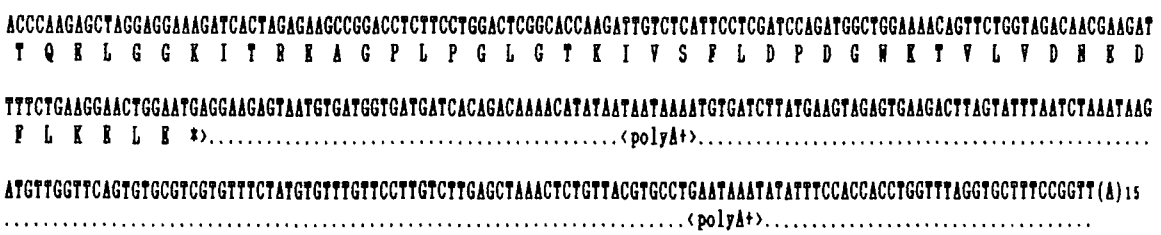

ACCCAAGAGCTAGGAGGAAAGATCACTAGAGAAGCCGGACCTCTTCCTGGACTCGGCACCAAGATTGTCTCATTCCTCGATCCAGATGGCTGGAAAACAGTTCTGGTAGACAACGAAGAT
T Q E L G G K I T R E A G P L P G L G T K I V S F L D P D G W K T V L V D N E D
TTTCTGAAGGAACTGGAATGAGGAAGAGTAATGTGATGGTGATGATCACAGACAAAACATATAATAATAAAATGTGATCTTATGAAGTAGAGTGAAGACTTAGTATTTAATCTAAATAAG
F L K E L E *)
(polyA+)
ATGTTGGTTCAGTGTGCGTCGTGTTTCTATGTGTTTGTTCCTTGTCTTGAGCTAAACTCTGTTACGTGCCTGAATAAATATATTTCCACCACCTGGTTTAGGTGCTTTCCGGTT(A)15
(polyA+)

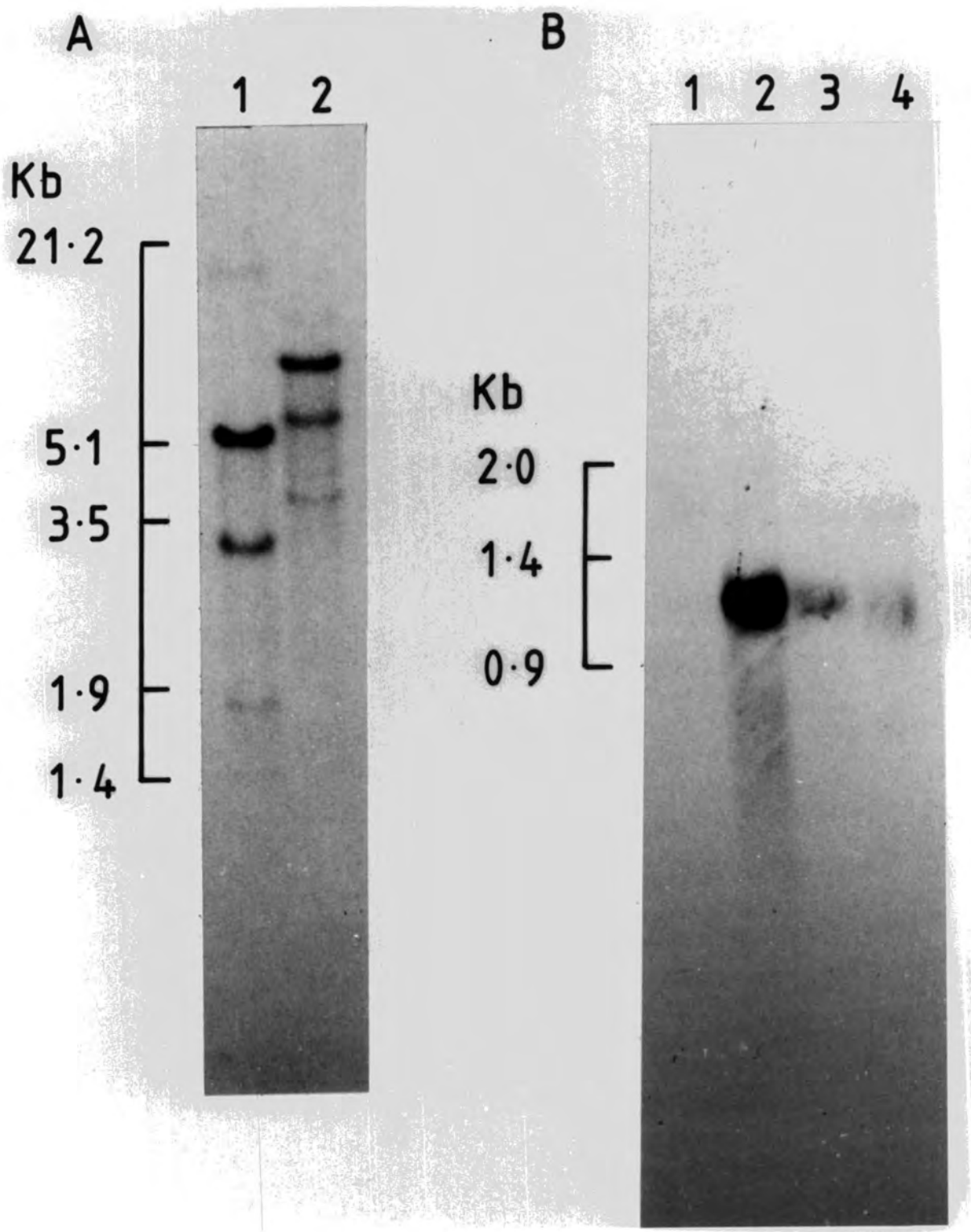
FIGURE 26

A

Autoradiograph of a genomic Southern hybridization of the 13G6(i) cDNA probe to restriction fragments of *Brassica oleracea* var. *gemmifera* S_5 genomic DNA. Genomic DNA samples (5μg per track) were restricted with: track 1, EcoRI; track 2, HindIII, prior to electrophoresis on a 0.7% agarose gel and Southern transfer. The hybridization was washed at high stringency (0.1 x SSC, 65°C).

B

Autoradiograph of a Northern hybridization of a probe of the 13G6(i) cDNA to total RNA extracts from *Brassica oleracea* var. *gemmifera* S_5 plants. The RNA samples (5μg per track) were from the following tissues: track 1, leaf; track 2. stigma (1 day prior to anthesis); track 3, style (1 day prior to anthesis); track 4, anthers (from flower buds of between 1 and 5mm in length).


A
1 2
Kb
21·2
5·1
3·5
1·9
1·4
B
1 2 3 4
Kb
2·0
1·4
0·9

4.7.1 Genomic Organization of 13G6(i)-Homologous Sequences

Figure 26A shows an autoradiograph of a Southern blot of *Brassica oleracea* genomic DNA probed with the 13G6(i) cDNA. Three bands are visible in tracks corresponding to both EcoR1 and HindIII restrictions of the genomic DNA.

4.7.2 Expression of Sequences Homologous to the 13G6(i) cDNA

mRNA transcripts homologous to cDNA 13G6(i) were detected by Northern blot hybridization in *Brassica* stigma, style and young anther tissue, but not in leaf tissue (Figure 26B). The signal obtained is much stronger in stigma than in style or anther samples and this is the case whether the hybridization is washed at low or high stringency, suggesting that all transcripts detected are highly homologous to the probe and that these are of greater percentage abundance in stigma tissue than in style or anther tissue.

The transcripts from style and stigma tissues are approximately 1200 b in length, whilst those detected from anther tissue are approximately 1100 b long. The slightly different transcript lengths may be explained in a number of ways: the genes transcribed in the various tissues may be separate homologous sequences utilizing different transcriptional start points; the transcripts (if products of a single gene) may be produced from different transcriptional start points, or be differently processed; tissue specific differential RNA splicing or the preferential use of the two poly(A) signals may be involved.

The length of the message to which clone 13G6(i) hybridizes in *Brassica* stigma RNA corresponds closely with the length of the cDNA, suggesting the cDNA to be at least close to full length.

CHAPTER FIVE

CONSPECTUS AND SUGGESTED FURTHER RESEARCH

5. CONSPECTUS AND SUGGESTED FURTHER RESEARCH

The analysis of a number of cDNA clones representing stigma-specific mRNA species has been described in the preceeding section, two of which encode members of a family of partially homologous glycoproteins which share a common pattern of tissue and temporal regulation of expression, and demonstrate considerable structural conservation. One of these (the cDNA from clone pBS2) shows many of the emergent characteristics of the SLSG sequences which have been cloned from a number of self-incompatible inbred lines of *Brassica*. It has, however, been stressed that this sequence is somewhat more diverged from the various putative SLSG sequences than are any of the remainder of them. Two separate classes of sequence, showing considerable homology and identical patterns of expression, are known from *Brassica* stigma cDNA libraries (SLSG and SLR1 cDNAs) and it is clearly possible that a third class may exist. Hence, clone pBS2 may contain a cDNA representing either the allele of the SLSG gene corresponding to the S_5 genotype of *B. oleracea*, or a third class of S-like gene.

Further work should be directed to identify the nature of the cDNA from pBS2, two experimental approaches being appropriate: RFLP analysis and gene-specific oligonucleotide probe hybridization studies. RFLP analysis should attempt to ascertain whether specific RFLP banding patterns from two homozygous self-incompatible parental lines of *Brassica* segregate with the SI phenotype in an F_2 generation. If they do, then the pBS2 cDNA may represent an allele of the SLSG gene. Genomic Southern blots of

restriction digests of a range of *Brassica* inbred lines should be probed with oligonucleotides derived from the central region of the pBS2 cDNA which shows lesser homology to other putative SLSG alleles. If oligonucleotide hybridization is specific to the S_5 genotype of *Brassica*, then this would support the contention that the pBS2 cDNA represents an SLSG allele.

Some work has been presented on cDNA clones other than those encoding 'S-like' glycoproteins. Of these, one is clearly a structural glycine-rich protein of probably no significance in the context of SI research. The remaining two classes of cDNA clones, exampled by clones 11H4 and 13G6(i) may have some significance in the reproductive biology of *Brassica*, though insufficient data has been amassed to make speculation on their possible roles worthwhile. Further work necessary to suggest the functions of these tissue-specifically expressed sequences might include more complete sequence analysis and an analysis of their tissue specificity of expression through *in situ* hybridization, or immunocytochemical localization of gene-products using antibodies raised against fusion proteins.

REFERENCES

Albersheim P, Anderson-Prouty AJ (1975) Carbohydrate, proteins, cell surfaces and the biochemistry of pathogenesis. Ann Rev Plant Physiol 26, 31-52

Amrhien N (1977) The current status of cyclic AMP in higher plants. Ann Rev Plant Physiol 28, 123-132

Anderson MA, Cornish EC, Mau S-L, Williams EG, Hoggart R, Atkinson A, Bonig I, Grego B, Simpson R, Roche PJ, Haley JD, Penschow JD, Niall HG, Tregear GW, Coghlan JP, Crawford, RJ, Clarke AE (1986) Cloning of a cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. Nature 321, 38-44

Aviv H, Leder P (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidilic acid-cellulose. Proc Natl Acad Sci USA 69, 1408-1412

Bennet MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Phil T Roy B 274, 227-274

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl Acid R 7, 1513-1523

Brewbaker JL (1957) Pollen cytology and incompatibility systems in plants. J Hered 48, 217-277

Charlesworth D (1988) Evolution of homomorphic sporophytic self-incompatibility. Heredity 60, 445-453

Clarke AE, Anderson MA, Bacic T, Harris PJ, Mao S-L (1985) Molecular basis of cell recognition during fertilization in higher plants. J Cell Sci Suppl 2, 261-285.

Condit CM, Meagher RB (1987) Expression of a gene encoding a glycine-rich protein in *Petunia*. Mol Cell Biol 7 (12), 4273-4279

Cornish EC, Pettitt JM, Bonig I, Clarke AE (1987) Developmentally controlled expression of a gene associated with self-incompatibility in *Nicotiana alata*. Nature 326, 99-102

Crowe LK (1964) The evolution of outbreeding in plants I-the angiosperms. Heredity 19, 437-457

Cuellar R (1982) Ph D Thesis. Stanford University, USA

Darwin C. (1877) The Different Forms of Flowers on Plants of the Same Species. John Murray: London

Day PR (1974) Genetics of Host-Parasite Interaction. Freeman and Co: San Francisco, USA

Dickinson HG, Lewis D (1973) The formation of the tryphine coating the pollen grains of *Raphanus*, and its properties relating to the self-incompatibility system. Proc R Soc Lond B 184, 149-165

Donk JAW van der (1975) Recognition and gene expression during the incompatibility reaction in *Petunia* Hybrida L. Mol Gen Genet 141, 305-316

East EM (1940) The distribution of self-sterility in flowering plants. Proc Am. Phil. Soc. USA 82, 449-518

East EM, Mangelsdorf AJ (1925) A new interpretation of the hereditary behaviour of self-sterile plants. Proc Natl Acad Sci (Wash) 11, 166-177

Ebert PR, Anderson MA, Bernatzki R, Altschuler M, Clarke AE (1989) Genetic polymorphism of self-inconpatibility in flowering plants. Cell 56, 255-262

Elleman CJ, Dickinson HG (1986) Pollen-stigma interactions in *Brassica* IV- structural reorganisation in the pollen grains during hydration. J Cell Sci 80, 141-157

Ferrari TE, Lee SS, Wallace DH (1981) Biochemistry and physiology of recognition in pollen-stigma interactions. Phytpathology 71, 752-755

Fryxell PA (1957) Mode of reproduction of higher plants. Botanical Reviews 23, 135-233

Gibbs P (1986) Do homomorphic and heteromorphic self-incompatibility systems have the same sporophytic mechanism Plant Syst Evol 154, 285-323

Graham DE (1978) The isolation of high molecular weight DNA from whole organisms and large tissue masses. Anal Biochem 85, 609-613

Gubler U, Hoffman BJ (1983) A simple and very efficient method for generating cDNA libraries. Gene 25, 263-269

Hanahan D (1985) Techniques for transformation of E. coli. In: DNA Cloning Vol 1- A Practial Approach, (ed Glover DM). IRL Press : Oxford

Heslop-Harrison J (1968) Ribosome sites and S-gene action. Nature 218, 90-91

Heywood VH (1978) Flowering Plants of the World. Oxford University Press: Oxford

Hinata K, Nishio T, Kimura J (1982) Comparative studies on S-glycoproteins purified from different S-genotypes in self-incompatible *Brassica* species II- immunological specificities. Genetics 100, 649-657

Hodgekin T, Lyon GD (1984) Pollen germination inhibitors in extracts of *Brassica oleracea* L. stigmas. New Phytol 96, 293-298.

Hoggart R (1984) Ph D Thesis. University of Melbourne, Australia

Isogai A, Takayama S, Shiozawa H, Tsukamoto C, Kanbara T, Hinata K, Okazaki K, Suzuki A (1988) Existence of a common glycoprotein homologous to S-glycoproteins in two self-incompatible homozygotes of *Brassica campestris*. Plant Cell Physiol 29 (8), 1331-1336

Katsumata T, Takahashi N, Ejiri SI (1978) Changes of cAMP level and adenylate cyclase activity during germination of pine pollen. Agric Biol Chem 42, 2161-2162

Keller B, Sauer N, Lamb CJ (1988) Glycine-rich cell wall proteins in bean: gene structure and association of the protein with the vascular system. EMBO J 7 (12), 3625-3633

Keller B, Templeton MD, Lamb CJ (1989) Specific localization of a plant cell wall glycine-rich protein in protoxylem cells of the vascular system. Proc Natl Acad Sci USA 86, 1529-1533

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132

Lalonde BA, Nasrallah ME, Dwyer KG, Chen C-H, Barlow B, Nasrallah JB (1989) A highly conserved *Brassica* gene with homology to the S-locus-specific glycoprotein structural gene. The Plant Cell 1, 249-258

Lawrence MJ, Marshall DF, Curtis VE, Fearson CH (1985) Gametophytic self-incompatibility re-examined: a reply. Heredity 54, 131-138

Lewis D (1952) Serological reactions of pollen incompatibility substances. Proc R Soc Lond B 140, 127-135

Lewis D (1964) A protein dimer hypothesis on incompatibility. Proceedings of the 11th International Congress on Genetics, The Hague, NL. In: Genetics Today (ed Geerts SJ) 3, 656-663

Lewis D, Verma SC, Zuberi MI (1988) Gametophytic-sporophytic incompatibility in the Cruciferae- *Raphanus* sativus. Heredity 61, 255-366

Logemann J, Schell J, Willmitzer L (1987) Improved method for the isolation of RNA from plant tissues. Anal Biochem 163, 16-20

Lundqvist A (1964) The nature of the two-loci incompatibility system in grasses IV- interaction between the loci in relation to pseudo-compatibility in *Festuca pratensis*. Hereditas 52, 221-234

Malic CP, Chhabra N, Vermani S (1976) Cyclic AMP-induced elongation of pollen tubes in *Tradescantia paludosa*. Biochem Physiol Phlanz 169, 311-315

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, New York

Marsluff WF. Huang RC (1984) Transcription of RNA in isolated nuclei. In: Transcription and Translation: A Practical Approach (eds: Hames BD, Higgins SJ). IRL Press: Oxford

Mason PJ, Williams JG (1985) hybridization analysis of DNA. In: Nucleic Acid hybridization- A Practical Approach (eds Hames BD, Higgins SJ). IRL Press: Oxford

Messing J (1983) New M13 vectors for cloning. Methods Enzymol 101, 20-78

McMaster GK, Carmichael GG (1977) Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc Natl Acad Sci USA 74, 4835-4838

Mulcahy DL, Mulcahy GB (1983) Gametophytic self-incompatibility re-examined. Science 20, 1247-1252

Mulcahy DL, Mulcahy GB (1988) Tests of the heterosis model. Sex Plant Reprod 1, 32-35

Nasrallah JB, Doney RC, Nasrallah ME (1985a) Biosynthesis of glycoproteins involved in the pollen-stigma interaction of incompatibility in developing flowers of *Brassica oleracea* L. Planta 165, 100-107

Nasrallah JB, Kao T-H, Chen C-H, Goldberg ML, Nasrallah ME (1987) Amino acid sequence of glycoproteins encoded by three alleles of the S-locus in *Brassica oleracea*. Nature 226, 617-619

Nasrallah JB, Kao T-H, Goldberg ML, Nasrallah ME (1985b) A cDNA clone encoding an S-locus-specific glycoprotein from *Brassica oleracea*. Nature 318, 263-267

Nasrallah JB, Yu S-M, Nasrallah ME (1988) Self-incompatibility genes of *Brassica oleracea*. Expression, isolation and structure. Proc Natl Acad Sci USA 85, 5551-5555

Nettancourt D de (1977) Incompatibility in angiosperms. Monographs in Theoretical and Applied Genetics 3. Springer-Verlag: Berlin

Ockendon DJ (1975) Dominance relationships between S-alleles in the stigmas of Brussels sprouts (*Brassica oleracea* var. *gemmifera*). Euphytica 24, 165-172

Perler F, Efstratiadis A, Lomedico P, Gilbert W, Kolander R, Dodgson J (1980) The evolution of genes: the chicken preproinsulin gene. Cell 20, 555-566

Polya GM, Mioncci V, Rae AL, Harris PJ, Clarke AE (1986) Ca^{2+}-dependent protein phosphorylation in germinating pollen of *Nicotiana alata*, an ornamental tobbaco. Physiol Plant 67, 151-157

Reddy ASN, Poovaiah BW (1977) Accumulation of a glycine-rich protein in auxin-deprived strawberry fruits. Biochem Biophys Res Commun 147 (3), 885-891

Rendle AB (1967) The Classification of Flowering Plants Vol 2. Cambrige University Press: Cambridge

Roberts IN, Harrod G, Dickinson HG (1984) Pollen-stigma interactions in *Brassica* II- the fate of stigma surface proteins following pollination and their role in the self-incompatibility response. J Cell Sci 66, 255-264

Rougier M, Jnoud N, Dumas C (1988) Localization of adenylate cyclase activity in *Populus*: its relation to pollen-pistil recognition and incompatibility. Sex Plant Reprod 1, 140-149

Sanger F, Coulson AR, Barrel BG, Smith AJH, Rose B (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol 143, 161-178

Sarkar RH, Elleman CJ, Dickinson HG (1988) Control of pollen hydration in *Brassica* requires continued protein synthesis, and glycosylation is necessary for intraspecific incompatibility. Proc Natl Acad Sci USA 85, 4340-4344

Southern E (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98, 503-517

Stead AD, Roberts IN, Dickinson HG (1979) Pollen-pistil interactions in *Brassica oleracea*: events prior to pollen germination. Planta 146, 211-216

Takayama S, Isogai A, Tsukamoto C, Ueda Y, Hinata K, Okazaki K, Koseki K, Suzuki A (1986) Agric Biol Chem 50 (6), 1673-1676

Takayama S, Isogai A, Tsukamoto C, Ueda Y, Hinata K, Okazaki K, Suzuki A (1987) Sequences of S-glycoproteins, products of the *Brassica campestris* self-incompatibility locus. Nature 326, 102-105

Thomas PS (1980) hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc Natl Acad Sci USA 77, 5202-5205

Thomas PS, Farquar MN (1978) Specific measurement of DNA in nuclei and nucleic acids using diaminobenzoic acid. Anal Biochem 89, 35-44

Trick M and Flavell RB (1988) Closely related *Brassica* S-alleles and the determination of allelic specificity. Heredity 61, 287

Trick M, Flavell RB (1989) A homozygous S-genotype of *Brassica oleracea* expresses two S-like genes. Mol Gen Genet 218, 112-117

Wagh PV, Bahl OP (1981) Sugar residues on proteins. CRC Crit Rev Biochem 10, 307-377

Watson CJ, Jackson JF (1985) Synthesis of double-stranded cDNA for cloning. In: DNA Cloning Vol I- A Practical Approach (ed Glover DM). IRL Press: Oxford

Whitehouse HLK (1950) Multiple-allelomorph incompatibility of pollen and style in the evolution of the angiosperms. Ann Bot 14, 198-216

Woods D (1984) Oligonucleotide screening of cDNA libraries. Focus 6 (3), 1-3

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains. Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33 (1), 103-119

Zuberi MI, Lewis D (1988) Gametophytic-sporophytic incompatibility in the Cruciferae- *Brassica campestris*. Heredity 61, 367-377

Mol Gen Genet (1990) 220:409–413

MGG

A cDNA encoding an S-locus specific glycoprotein from *Brassica oleracea* plants containing the S_5 self-incompatibility allele

Charles P. Scutt, Phillip J. Gates, John A. Gatehouse, Donald Boulter, and Ronald R.D. Croy
Plant Molecular Biology Group, Department of Biological Sciences, University of Durham, South Road, Durham DH1 3LE, UK

Summary. A cDNA sequence homologous to the *Brassica* self-incompatibility locus specific glycoprotein (SLSG) sequence was isolated from stigmas of *B. oleracea* plants homozygous for the S_5 allele. The nucleotide sequence of this cDNA was obtained and compared with the S_6 allelic form of the SLSG. Evidence is presented which indicates that this sequence does not specify the self-incompatibility response of pollen.

Key words: Self-incompatibility – S_5 allele – SLSG – *Brassica oleracea*

Introduction

A large number of species from a diverse range of angiosperm families possess a self-incompatibility (SI) mechanism through which the post-pollination development of self pollen is inhibited and self-fertilization thus prevented. This is interpreted as an adaptive mechanism promoting heterozygosity and thereby retaining many recessive alleles within the gene pool. Such SI mechanisms have been hypothesized to have made a major contribution to the success of the angiosperms and to have been present from a very early stage in their evolution (Whitehouse 1950). However, the lack of homology between sequence data from different species in genes putatively involved in SI recognition (Nasrallah et al. 1985; Anderson et al. 1986) suggests a polyphyletic origin for at least some extant SI systems.

A relatively small number of SI systems have been studied intensively. Of these the *Brassica* SI system is probably the most highly investigated system of the sporophytic type. This term arises from the genetic control of SI in pollen, as in sporophytic SI systems it is the diploid parent plant (or sporophyte) which confers the SI recognition phenotype on the pollen grains. Pollen is incompatible with plants sharing either of the SI alleles (S-alleles) of its parent. In *Brassica*, over 50 allelic forms of the single S-locus, which controls the SI recognition type of pollen and stigma, have been identified.

Molecular studies of SI in *Brassica* species have succeeded in characterising a number of allelic cDNA and genomic sequences which encode an SI locus specific glycoprotein (SLSG), synthesized in stigmas (Nasrallah et al. 1985, 1987, 1988). The mode of expression, linkage and putative glycoprotein product of the SLSG gene strongly suggest a role for it in encoding the female-expressed recognition component of the SI system, though definitive evidence for this remains to be established. In contrast no SI recognition component has yet been identified from *Brassica* pollen.

This paper describes the isolation and characterisation of a cDNA sequence for the S_5 allelic form of the SLSG gene expressed in stigmas of *B. oleracea*. A number of different fragments of genomic DNA are shown to bear varying degrees of homology to the SLSG sequence. Evidence is presented which suggests that neither the SLSG gene nor any closely related (cross-hybridising) sequence is transcribed in anther tissue during stages of development at which a sequence specifying the SI reaction of pollen would be expected to be expressed. This expectation is based on the widely accepted theory of Heslop-Harrison (1968) which proposes the site of synthesis of the male determinant of SI in sporophytic systems to be the tapetum surrounding the cells of the male germ line within the developing anthers.

Materials and methods

Plant material. Homozygous inbred lines of *B. oleracea* var. *gemmifera*, S_5, and *B. alboglabra*, (Bailey), S_5, (a rapid cycling variety) were a generous gift from Dr. D. Ockendon of the Institute of Horticultural Research, Wellesbourne, Warwick, UK. These lines were perpetuated in Durham by self pollination of immature flower buds at approximately 2 days before anthesis.

RNA extraction. RNA was extracted from plant tissues by the method of Logemann et al. (1987) and poly(A)$^+$ RNA was selected by oligo(dT) cellulose affinity chromatography according to Evans et al. (1979). Stigma poly(A)$^+$ RNA was prepared from stigma tissue harvested from flower buds of *B. oleracea* var. *gemmifera* S_5 plants between 2 and 0 days before anthesis.

Construction of cDNA library. Stigma cDNA was synthesized from poly(A)$^+$ RNA using a cDNA synthesis kit (BCL, UK) and used according to the supplier's instructions. The double-stranded cDNA was ligated into the

Abbreviations: SDS, sodium dodecyl sulphate; PVP, polyvinylpyrrolidone; BSA, bovine serum albumin; SLSG, self-incompatibility locus specific glycoprotein

Offprint requests to: R.R.D. Croy

*Eco*RI site of phosphatase-treated pUC18 vector, following the addition of *Eco*RI synthetic linkers (Croy et al. 1982). Transformation of *Escherichia coli* strain DH5-α competent cells (library grade cells, Gibco-BRL) resulted in a library of approximately 20000 recombinant clones.

Screening of the library for SLSG sequences. The library was screened with a 30-mer oligonucleotide probe synthesised on an Applied Biosystems 381A DNA Synthesizer and corresponding to nucleotides +1096 to +1126 of the published S_6 cDNA sequence (5'-TGC CTT AGC GAT TGT AAT TGT ACT GCT TTT-3') (Nasrallah et al. 1985). Oligonucleotide probe labelling and colony screening on duplicate nitrocellulose filters were carried out as described by Woods (1984). Hybridisation was performed at 35° C for 15 h. Final washing of filters was carried out at low stringency, to allow for the detection of imperfectly matched hybrids (6 × SSC, 0.1% w/w SDS, 45° C for 1 h). (20 × SSC is 0.3 M NaCl, 0.3 M sodium citrated, pH 7.0.) The screening produced three positive clones designated pBS1, pBS2 and pBS3, containing cDNA inserts of 400, 1600 and 500 bp respectively. The longest cDNA was selected for further work.

M13 subcloning and sequence analysis. Suitable restriction fragments from the selected cDNA sequence were generated and subcloned into M13 mp18 and mp19 vectors for sequencing by the dideoxy chain termination method (Sanger et al. 1980). Both strands of the cDNA were completely sequenced.

Southern hybridisation. Genomic DNA was extracted from young leaf tissue as described by Graham (1978) and was purified once on caesium chloride/ethidium bromide gradients. Aliquots of DNA (5 µg) were digested to completion with restriction endonuclease and electrophoresed on neutral agarose gels along with amounts of the SLSG cDNA equivalent to 1 to 4 gene copies in the genomic DNA. Gene copy equivalent calculations were based on a haploid genome size of 8.6×10^8 bp for *B. oleracea* (Bennet and Smith 1976). Gels were blotted and hybridised to the SLSG cDNA excised from pBS2, and labelled with α-[^{32}P]dCTP to a specific activity of 2×10^8 cpm/µg using random oligonucleotide priming according to Feinberg and Vogelstein (1983). Hybridisation was performed in 5 × SSC, 5 × Denhardts, 0.1% (w/v) SDS and 100 µg/ml denatured, sonicated salmon sperm DNA, as detailed by Maniatis et al. (1982). (10 × Denhardts is 0.2% (w/v) Ficoll, 0.2% (w/v) PVP, 0.2% (w/v) BSA.) Filters were washed initially to low stringency (2 × SSC, 0.1% (w/v) SDS, 65° C, 30 min) exposed, and then washed to high stringency (0.1 × SSC, 0.1% (w/v) SDS, 65° C, 15 min) and re-exposed. Filters were exposed to pre-sensitized X-ray film for appropriate times, using Du Pont Cronex intensifying screens.

Northern hybridisation. Total RNA preparations were denatured by treatment with deionized glyoxal and electrophoresed on a high gelling temperature agarose gel (Thomas 1980). Duplicate sets of RNA samples were analysed on the same gel; half of each set was stained with acridine orange to locate the RNA band positions and the other half was blotted onto nitrocellulose according to Thomas (1980). Hybridisation was performed at 42° C in a solution containing 50% (v/v) deionized formamide, 5 × SSPE, 5 × Denhardts, 0.1% (w/v) SDS, 100 µg/ml denatured, sonicated salmon sperm DNA. (20 × SSPE is 0.3 M NaCl, 0.02 M EDTA, 0.2 M NaH_2PO_4, pH 7.4.) Final washing of filters was to low stringency (1 × SSPE, 0.1% (w/v) SDS, 42° C, 1 h).

Results

The S_5 cDNA sequence

Screening of the stigma cDNA library revealed three independent clones hybridising to the oligonucleotide probe. Cross hybridisation experiments demonstrated close homology between the three clones (results not presented). The longest of these cDNA molecules was fully sequenced and contains an open reading frame of 1284 bases (427 amino acids), a non-coding 3' region of 178 bases and a poly(A) tail (Fig. 1). There is no in-frame start codon, though the sequence maintains its homology with other SLSG sequences for some distance 5' to the N-terminus of the mature protein into the putative signal peptide region. The S_5 cDNA sequence shows an anomalous extreme 5' end, which bears a poly(dT) 'tail'. This suggests that the cloned sequence represents the SLSG cDNA, minus the initiation sequence, ligated at its 5' end to a short stretch of a second cDNA. Two sequences conforming to the polyadenylation signal motif occur within the 3' non-coding region of the cDNA.

The S_5 stigma cDNA sequence represents a form of the SLSG gene. It has close homology to the S_6-SLSG cDNA sequence reported by Nasrallah et al. (1985), as shown in Fig. 1. Overall the S_5 and S_6 sequences show 71% homology at the DNA level and 64% homology at the amino acid level. The 30-mer oligonucleotide used to probe the stigma library selectively hybridised to the S_5 SLSG cDNA despite five mismatches occurring over the 30 bp region from which the probe was derived. The oligonucleotide probe could be removed from the S_5 cDNA on blots by washing at increasing stringency such as temperatures of between 55° C and 60° C in 5 × SSC (results not presented).

Genomic organization of S-related sequences

Hybridisation of the S_5 cDNA to genomic blots (Fig. 2) indicated a number of genomic sequences with homology to the SLSG gene. Hybridization of the probe to fragments of lower homology was reduced by high stringency washing (Fig. 2). The gene copy reconstructions showed no more than one highly homologous sequence present on any single restriction fragment of genomic DNA. This indicates that the gene sequence is present in only 2 or 3 copies per haploid genome. Comparison of genomic banding patterns between *B. oleracea* S_5 and *B. alboglabra* S_5 suggests a similar organisation of S-related sequences within their respective genomes (Fig. 2).

Expression of the SLSG gene

Northern hybridisation shows the SLSG cDNA clone to represent a stigma specific mRNA of approximately 1600 bases (Fig. 3). The RNA, while expressed to high level in stigma tissue, was not detectable in leaf tissue. Furthermore no homologous RNA could be detected in any of

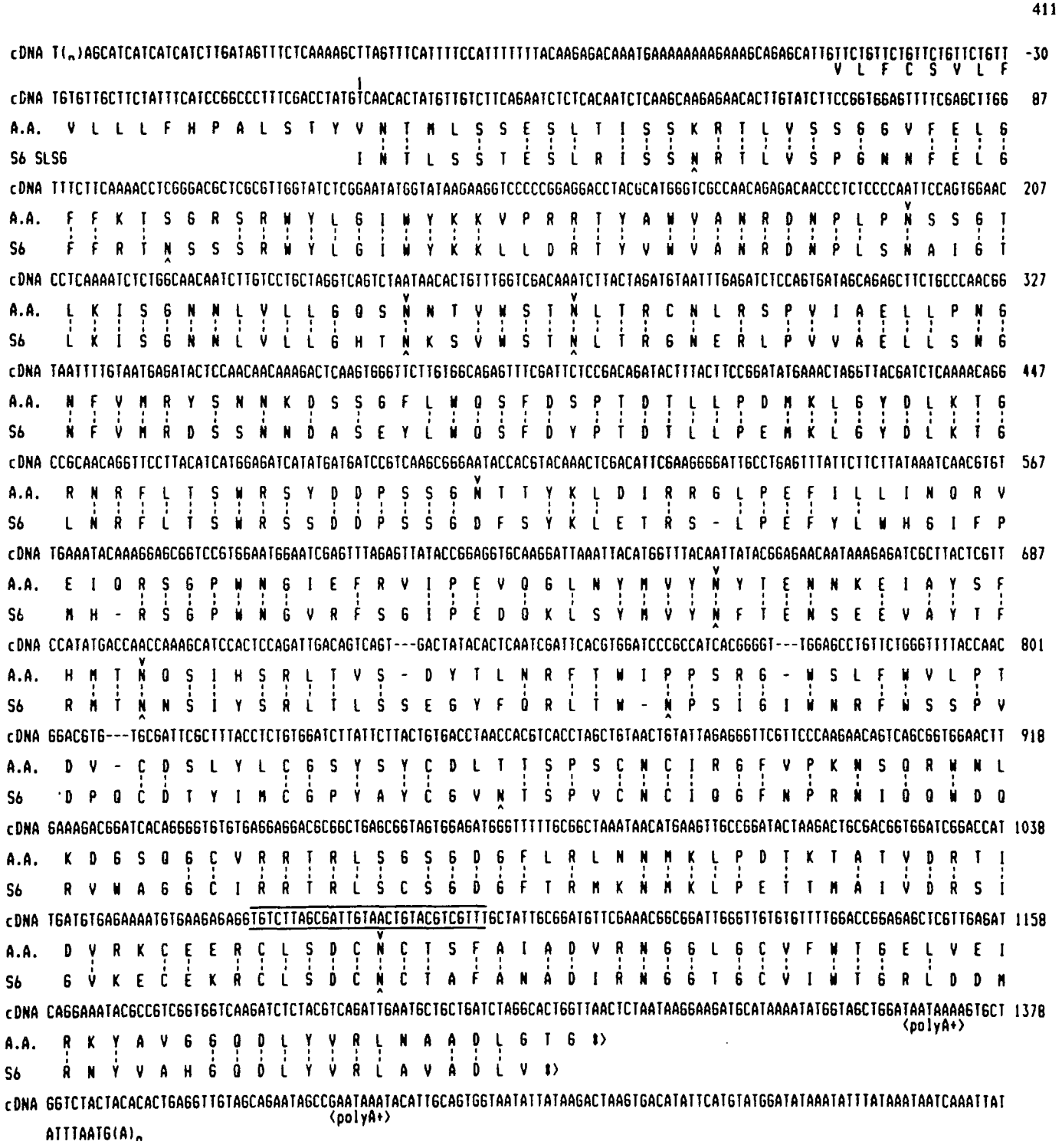


Fig. 1. Nucleotide sequence of the S_5 SLSG cDNA. The amino acid sequence derived from the DNA is presented below the sequence and compared with that of the S_6 SLSG amino acid sequence from Nasrallah et al. (1985). The putative N-glycosylation signals are indicated by *arrowheads*. The region of the S_5 cDNA hybridising to the S_6 oligonucleotide probe is *over-* and *underlined*. The position of the first codon of the mature protein is numbered 1

the tissues of the anther at stages of development correlating with flower bud lengths of between 1 and 5 mm. Fluorescence microscopy (results not presented) indicates tetrads to be present within anthers of 1 mm long buds and the tapetal tissue to have degraded within the anthers of 5 mm long buds. A much longer exposure of the Northern hybridisation blot (not presented) similarly failed to reveal any band hybridising to the SLSG probe at any stage of anther development examined. It is estimated that the technique employed is capable of detecting a level of expression of the SLSG gene three orders of magnitude less than that encountered in stigma tissue 1 day prior to anthesis.

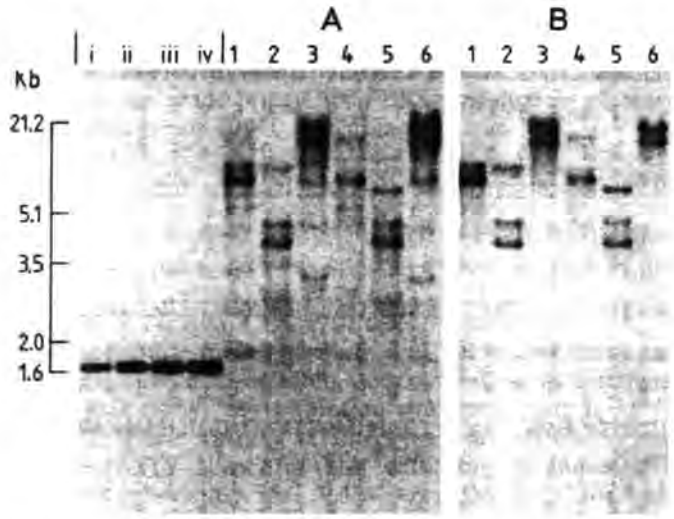


Fig. 2A and B. Southern blots showing hybridisation of the S_5 cDNA to genomic restriction fragments. Blot exposed (**A**) after low stringency wash and (**B**) after high stringency wash. Lanes 1, 2 and 3, *Brassica oleracea* var. *gemmifera* DNA (5 μg per lane); lanes 4, 5 and 6, *B. alboglabra* DNA (5 μg per lane). Lanes 1 and 4, *Eco*RI digests; lanes 2 and 5, *Hin*dIII digests; lanes 3 and 6, *Kpn*I digests. Lanes i–iv contain 1, 2, 3 and 4 gene copy equivalents

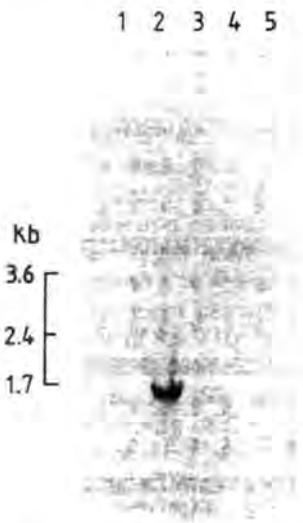


Fig. 3. RNA blots showing hybridisation of the S_5 SLSG cDNA to *Brassica alboglabra* RNA extracts from (1) leaf, (2) stigma (1 day prior to anthesis), (3–5) anthers from buds of 1–2 mm, 2–3 mm and 3–5 mm respectively (5 μg RNA per lane)

Discussion

This paper contributes a new *Brassica* SLSG sequence thus enabling further comparison between allelic forms of the protein and its gene. The sequence was isolated from a stigma cDNA library using a synthetic oligonucleotide corresponding to a sequence highly conserved in other S-alleles of the SLSG. The S_5 sequence appears to fit within the pattern of homology exhibited between the published *Brassica* SLSG sequences (Fig. 1) (Nasrallah et al. 1987; Takayama et al. 1987; Trick and Flavell 1989). It would appear from the sequence comparison that structural features of the SLSG in the SI reaction might include its pattern of glycosylation and/or the sequence of its more variable central region. It must, however, be stressed in conjectures relating to the variable structure of the SLSG molecule that its role as the female-expressed recognition component in SI has yet to be conclusively demonstrated.

An expressed sequence with nucleic acid homology of 59% to the S_5 SLSG cDNA has been reported as a cDNA (Lalonde et al. 1989; Trick and Flavell 1989) and as protein sequence data (Isogai et al. 1988). Differential screening of the cDNA library used in the present work with total cDNA probes prepared from both stigma and leaf mRNA has resulted in the selection of another full-length stigma specific cDNA (sequence not presented) which encodes this SLSG-like sequence. It is identical in coding sequence to that reported by Lalonde et al. (1989), and termed the SLR gene, except for a single nucleotide substitution (G for A at base 1307 of the cDNA) which results in the substitution of arginine for lysine at amino acid residue 407 of the predicted sequence of the mature protein. It is of interest that the 3′ non-coding regions of the SLR cDNAs cloned from S_5 and S_{22} plants diverge markedly after nucleic acid residue 1345. The very high level of conservation of the SLR sequence between *Brassica* plants expressing different SI phenotypes and the demonstration (Lalonde et al. 1989) that the SLR gene is not closely linked to S show that the SLR gene cannot be responsible for the specification of SI in *Brassica*.

The Southern blot hybridisation data presented (Fig. 2) confirm the assertion made by Nasrallah et al. (1988) that a number of SLSG-related sequences, perhaps as many as 11, reside within the *Brassica* genome. The approach taken in this experiment demonstrates that these sequences cannot entirely be discriminated between by adjusting the stringency of hybridisation or washing conditions.

Classical genetic studies on SI in *Brassica* have led, in describing the S-locus, to the concept of a single gene sequence, expressed in both male and female tissues, which specifies SI: "the entire chain of rejection events is initiated by a recognition reaction which occurs between identical incompatibility substances on the pollen exine and the stigmatic pellicle" (de Nettancourt 1977). However, no sequence closely homologous to the SLSG sequence could be detected in anther tissue between developmental stages corresponding to the separation of the meiotic products and the autolysis of the tapetum. It is possible that the SLSG gene is transcribed in the tapetum at a very early stage in development and therefore missed in the Northern hybridisation presented. However, this seems highly unlikely as the male-expressed S-gene product should be amongst the last of the substances of tapetal origin to be exported to the developing pollen grains; various lines of evidence suggest this male specified product resides in the superficial coating of the mature pollen. The apparent lack of expression of the SLSG gene in anther tissue reported in the present work may be tentatively interpreted as evidence for a different, though genetically linked, sequence, responsible for the specification of SI in pollen grains.

Acknowledgements. The authors would like to acknowledge the receipt of an SERC studentship to C.S., and the financial assistance of the University of Durham Research Fund for the provision of oligonucleotide synthesis facilities. The technical assistance of Mr. P. Preston is gratefully acknowledged.

References

Anderson MA, Cornish EC, Mau S-L, Williams EG, Hoggart R, Atkinson A, Bonig I, Grego B, Simpson R, Roche PJ, Haley JD, Penscho JD, Niall HG, Tregear GW, Coghlan JP, Crawford RJ, Clarke AE (1986) Cloning of a cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. Nature 321:38–44

Bennet MD, Smith JB (1976) Nuclear DNA amounts in Angiosperms. Philos Trans R Soc Lond 274:227–274

Croy RRD, Lycett GW, Gatehouse JA, Yarwood JN, Boulter D (1982) Cloning and analysis of cDNAs encoding plant storage protein precursors. Nature 295:76–79

De Nettancourt D (1977) Incompatibility in Angiosperms. Springer, Berlin

Evans IM, Croy RRD, Hutchinson P, Boulter D, Payne P, Gordon ME (1979) Cell free synthesis of some storage protein subunits by polyribosomes and RNA isolated from developing seeds of pea (*Pisum sativum* L.). Planta 144:455–462

Feinberg A, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Graham DE (1978) The isolation of high molecular weight DNA from whole organisms or large tissue masses. Anal Biochem 85:609–613

Heslop-Harrison J (1968) Ribosome site and S-gene action. Nature 218:90–91

Isogai A, Takayama S, Shiozawa H, Tsukamoto C, Kanbara T, Hinata K, Okazaki K, Suzuki A (1988) Existance of a common glycoprotein homologous to S-glycoproteins in two self-incompatible homozygotes of Brassica campestris. Plant Cell Physiol 29:1331–1336

Lalonde BA, Nasrallah ME, Dwyer KG, Chen C-H, Barlow B, Nasrallah JB (1989) A highly conserved Brassica gene with homology to the S-locus-specific glycoprotein structural gene. Plant Cell 1:249–258

Logemann J, Schell J, Willmitzer L (1987) Improved method for the isolation of RNA from plant tissues. Anal Biochem 163:16–20

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: A laboratory manual. Cold Spring Harbour Laboratory, New York

Nasrallah JB, Kao T-H, Goldberg ML, Nasrallah ME (1985) A cDNA clone encoding an S-locus specific glycoprotein from *Brassica oleracea*. Nature 318:263–267

Nasrallah JB, Kao T-H, Chen C-H, Goldberg ML, Nasrallah ME (1987) Amino-acid sequence of glycoproteins encoded by three alleles of the S-locus of *Brassica oleracea*. Nature 326:617–619

Nasrallah JB, Yu S-M, Nasrallah ME (1988) Self-incompatibility genes of *Brassica oleracea*: Expression, isolation and structure. Proc Natl Acad Sci USA 85:5551–5555

Sanger F, Coulson AR, Barrel BG, Smith AJH, Rose B (1980) Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol 143:161–178

Takayama S, Isogai A, Tsukamoto C, Ueda Y, Hinata K, Okasaki K, Susuki A (1987) Sequences of S-glycoproteins, products of the *Brassica campestris* self-incompatibility locus. Nature 326:102–105

Thomas PS (1980) Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc Natl Acad Sci USA 77:5202–5205

Trick M, Flavell RB (1989) A homozygous S genotype of Brassica oleracea expresses two S-like genes. Mol Gen Genet 218:112–117

Whitehouse HLK (1950) Multiple-allelomorph incompatibility of pollen and style in the evolution of the angiosperms. Ann Bot 14:198–216

Woods D (1984) Oligonucleotide screening of cDNA libraries. Focus 6(3):1–3

Communicated by H. Saedler

Received March 20, 1989