Sequence Variability of Three Alleles of the Self-Incompatibility Gene of *Nicotiana alata*

Marilyn A. Anderson,* Geoffrey I. McFadden,* Robert Bernatzky,* Angela Atkinson,* Timothy Orpin,* Helen Dedman,* Geoffrey Tregear,*b Ross Fernley,* and Adrienne E. Clarke*†

*Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Parkville, Victoria 3052, Australia

bHoward Florey Institute of Experimental Physiology and Medicine, University of Melbourne, Parkville, Victoria 3052, Australia

Three alleles of the self-incompatibility gene of *Nicotiana alata* have been cloned and sequenced. A comparison of the sequences shows a surprisingly low level of homology (56%) and the presence of defined regions of homology and variability. The homologous regions include the N-terminal sequence, most of the cysteine residues and glycosylation sites, as well as other blocks throughout the sequence. We interpret these conserved regions as "framework" and nonconserved regions as "hypervariable," following the terminology used to describe analogous regions in the IgG supergene family. The low level of overall homology forms the basis of a general method for isolating S-allele-specific cDNAs. Allele-specific antibodies can be generated using synthetic peptides corresponding to one of the variable regions. When applied to sections of the pistil, these antibodies label the intercellular matrix in the stigma and transmitting tissue of the style and the cell walls in the epidermis of the placenta. HindIII digestion of genomic DNA generates a characteristic pattern of S-gene fragments for each genotype. These restriction fragment length polymorphisms can be used to assign S-genotype to progeny arising from breeding experiments.

INTRODUCTION

The evolutionary success of flowering plants is attributed in part to the development of self-incompatibility, which acts to prevent self-fertilization and promote outcrossing (see de Nettancourt, 1977; Harris et al., 1984; Cornish, Pettitt, and Clarke, 1988, for reviews). In many plants with a genetically controlled self-incompatibility (SI) system, the specificity of the SI reaction is controlled by a single gene (S-gene), which has multiple alleles. Pollen is rejected if the same S-allele is expressed by both the pollen and the pistil. Self-incompatibility is divided into two main types: sporophytic, in which the pollen has the same S-phenotype as the parent, and gametophytic, in which pollen phenotype is determined by the haploid genotype of the pollen. S-Allele-associated glycoproteins have been identified in the styles of a number of plants which exhibit either gametophytic or sporophytic incompatibility. cDNAs encoding these glycoproteins have been isolated from *Nicotiana alata* (gametophytic SI, Anderson et al., 1986) and *Brassica oleracea* (sporophytic SI, Nasrallah et al., 1985, 1987).

In this paper, we describe the sequences of three alleles of the self-incompatibility gene of *N. alata*.

† To whom correspondence should be addressed.
RNA species of about 650 nucleotides was also evident in the S3-genotype and was detected with both the S2-cDNA probe (see Figure 1B) and the S3-cDNA probe (see Figure 1C). The S2-cDNA shares 64% homology with the S3-cDNA at the amino acid level (Figure 2), and the aminoterminal sequence matches that of the purified S2-glycoprotein derived by N-terminal amino acid sequencing (Jahnen et al., 1989). The signal sequence of the S2-glycoprotein is incomplete because the longest S2-cDNA clone (740 bp) terminated at an EcoRI site in the region coding for amino acids at positions 25 to 27 (see Figure 2). The rest of the sequence was obtained by RNA sequencing (Hamlyn, Gait, and Milstein, 1981). The S2-cDNA clones bound most strongly to an RNA species of about 900 bases from plants bearing an S2-allele (see Figure 1C). The predicted amino acid sequence of the longest cDNA clone (803 bp) is shown in Figure 2. Overall, the homology at the amino acid level with the S2-glycoprotein is 62% and the predicted amino acid sequence of the S2-cDNA contains the aminoterminal sequence of the mature S2-associated glycoprotein (Jahnen et al., 1989). Alignment of the amino acid sequences derived from the S2-, S3-, and S6-cDNA sequences (see Figure 2) shows that there is about 56% homology overall. Although there are changes scattered throughout the sequence, the variation is mainly associated with four hypervariable regions (see Figure 2) which are predicted to be hydrophilic (see Figure 4A, regions A, B, C, and D). The conserved regions are associated with the amino terminus, the cysteine residues, the glycosylation sites, and the carboxy terminus. The 9 cysteine residues of the mature S2-glycoprotein are all conserved in the S3- and S6-glycoproteins, but the S2- and S6-glycoproteins have an additional cysteine residue at position 79. The S2-, S3-, and S6-glycoproteins contain 4, 5, and 4 consensus sequences for N-glycosylation, respectively. Four of the N-glycosylation sites are in conserved positions in the three alleles, and none are included in the hypervariable regions apart from the fifth site, unique to the S3-glycoprotein (see Figure 2, position 102).
DNA Gel Blot Analysis and Breeding Experiments

The question of whether these sequences represent allelic products of the same S-locus or products of distinct genes is addressed by DNA gel blot analyses and breeding experiments. Figure 3A shows a DNA gel blot analysis of a HindIII digest of S,S, S,S, S,S, and S,S genomic DNA using the S-cDNA as probe. The probe hybridized to single fragments from the S,S, S,S, and S,S genomic DNA and to two fragments from S,S genomic DNA. The S-S and S,cDNA probes revealed the same restriction fragments, except that each cDNA clone bound strongly to fragments from homologous DNA and weakly to fragments from heterologous DNA (see Figure 3A). The weak cross-hybridization is consistent with the variation distributed throughout the cDNA sequences.

Breeding experiments were performed to determine whether the restriction fragments identified by the three cDNAs segregated with S-genotype in the F1-generation. Two plants with no common S-alleles (S,S, S, and S,S) were crossed to produce 27 offspring, which were divided into incompatibility classes according to their ability to accept or reject S,S, S,S, or S,S-pollen. Four incompatibility classes S,S (7 plants), S,S (4 plants), S,S (8 plants), and S,S (7 plants) were obtained as predicted from the one locus hypothesis (de Nettancourt, 1977; Lawrence, Afzal, and Kenrick, 1978). A second population was produced by crossing two plants with one S-allele in common (S,S, S, and S,S) in the population; this population contained 21 plants, which fell into the two predicted self-incompatibility classes, S,S (14 plants) and S,S (7 plants). DNA gel blot analyses of these two populations (see Figures 3B and 3C) showed that, for each plant, the S-genotype assigned by breeding behavior was identical with that predicted by analysis of the restriction fragment length polymorphisms (RFLPs). Although the patterns described are for the major hybridizing bands, there are fainter bands that become apparent after longer exposure. These are indicated by arrows in Figures 3A and 3C.

Production of an Antibody Specific to the S-Glycoprotein

An antibody was raised to a synthetic peptide corresponding to a stretch of the major hypervariable region B, which is hydrophilic (see Figure 4A, Region B) and thus predicted to lie on the surface of the glycoprotein (Hopp and Woods, 1981). This antibody bound specifically to the S-glycoprotein in protein gel blots under conditions where there was no detectable binding to the other buffer-soluble style proteins or to other S-allele glycoproteins (see Figure 4B).

Immunolocalization of the S-Associated Glycoprotein

Light microscope localization of antibody binding is shown in Figure 5. When applied to longitudinal sections of the
pistil, the antibody bound primarily to the extracellular material in the stigma and transmitting tract of the style. In the ovary, the antibody was localized on the epidermis of the placenta and in a thin layer surrounding each ovule. Negligible binding occurred when the antibody to the synthetic S$_2$-peptide was omitted or replaced with an antibody directed to the synthetic arginine vasopressin-peptide (data not shown).

The electron micrographs of the style transmitting tissue in Figure 6 showed specific localization of the antibody to the extracellular matrix. There was a little binding to the wall of the transmitting tissue cells and very little binding

---

**Figure 4.** An Antibody Raised to a Peptide Sequence of One of the Hypervariable Regions of the S$_2$-Glycoprotein Binds Specifically to the S$_2$-Glycoprotein.

**A** Hydrophobicity plot of the deduced amino acid sequence of the S$_2$-glycoprotein. Values above the line denote hydrophobic regions and values below the line represent hydrophilic regions. The putative hydrophobic signal peptide (amino acids -22 to -1) is stippled. The potential N-glycosylation sites are marked with arrows. The regions of the S$_2$-glycoprotein with greatest sequence diversity from the S$_2$- and S$_3$-glycoproteins are shaded (A to D). Region B contains the peptide sequence used to generate the S$_2$-specific antibody. The hydrophobicity profile was generated using the predictive rules of Kyte and Doolittle (1982) and a span of 9 consecutive residues.

**B** Specificity of antibodies raised against the synthetic S$_2$-peptide.

(a) Protein gel blot of style extracts probed with antibody raised to the synthetic S$_2$-peptide. The antibody specifically binds to the 32-kD S$_2$-associated glycoprotein. There is no detectable binding to other style proteins or to S-glycoproteins of other genotypes.

(b) Silver-stained SDS-polyacrylamide gel (12.5%) displaying extracts from mature styles of various N. alata S-genotypes. The S$_1$, S$_2$, S$_3$, and S$_4$-allele-associated proteins are indicated by arrows. The S$_2$-associated protein is present in lower amounts than the other S$_2$-allele-associated glycoproteins and is often masked by a protein of the same mobility, which is present in all genotypes (Jahnen et al., 1989).

---

**Figure 5.** Binding of Anti-S$_2$-peptide Antibody to Ovary and Style.

(A) Longitudinal section of stigma labeled with anti-S$_2$-peptide antibody and immunofluorescent marker. The walls of the stigma cells, including the surface papillae, are labeled. The cortical tissues (CO) and epidermis of the style (arrows) are not labeled. (Scale bar = 100 μm.)

(B) Longitudinal section of the style showing cortical tissues (CO) and transmitting tract (TT). The intercellular matrix in the transmitting tract is labeled. (Scale bar = 100 μm.)

(C) Longitudinal section of locule showing labeling of placental epidermis (arrows) and the outer surface of the ovules (OV). (Scale bar = 50 μm.)
Figure 6. Electron Micrographs of Style and Ovary Showing Immunogold Localization of Anti-S$_2$-Peptide Antibody (Scale bars = 500 nm.)

(A) Transverse section of style through transmitting tract region labeled with anti-S$_2$-peptide antibody. The intercellular matrix (IM) is heavily labeled with immunogold markers. In addition, there is some labeling of the cytoplasm of the transmitting tract cells. Amyloplasts (AM), mitochondria (MI), and vacuoles (VA) are not labeled.

(B) Similar section to (A) labeled with control antibody. There is no specific labeling in any region.

(C) Longitudinal section through locule (LO) labeled with S$_2$-peptide antibody showing two cells of placental epidermis. The cell wall is heavily labeled.
to the transmitting tissue cells (see Figure 6A). The control anti-arginine vasopressin antibody did not label any structures in the style (see Figure 6B). In the locule, the labeling is concentrated over the primary wall of the placental epidermal cells (see Figure 6C).

**DISCUSSION**

A major finding of this study is the extent of variation between the alleles. Alignment of the amino acid sequences derived from the $S_2^-$, $S_3^-$, and $S_6^-$ cDNA sequences shows that, although there is more than 50% homology, there are also regions of considerable diversity. We describe these nonconserved regions as "hypervariable" and the conserved regions as "framework" regions following the terminology described by Wu and Kabat (1970) for immunoglobulin structure. This feature of hypervariable and framework sequences in related proteins is common to a number of animal proteins encoded by the immunoglobulin gene superfamily (Williams and Barclay, 1988). Each of these genes is concerned with some aspect of cell recognition, and this seems to be a characteristic feature of this class of molecules. The products of the $S$ alleles are presumed to have a recognition function, as the growth of pollen tubes bearing a particular allele is arrested within the style of plants bearing the same allele (de Nettancourt, 1977).

It is possible that the conserved cysteine residues act to form disulfide bonds and present the hypervariable regions in binding domains as for the immunoglobulin supergene family. The variation in the amino acid sequence of the $S$-allele products is associated with hydrophilic regions, which are predicted to be on the surface of the molecule (Hopp and Woods, 1981), and would be accessible for binding to appropriate receptors in the pollen or pollen tube. Antibodies raised to a synthetic peptide corresponding to the stretch of the major hypervariable region bound to the native $S_2$-glycoprotein in fresh and fixed tissue sections as well as in enzyme-linked immunosorbent assays (data not shown), suggesting that this peptide is exposed at the surface of the glycoprotein. The use of the synthetic peptide as an antigen obviated the difficulties previously encountered using either polyclonal or monoclonal antibodies raised to purified $S_2$-glycoprotein. These difficulties of cross-reactivity with other style glycoproteins were due to the glycosyl substituents on the $S_2$-glycoprotein, which have saccharide sequences common to many plant glycoproteins and which are immunodominant (Howlett and Clarke, 1981; Anderson, Sandrin, and Clarke, 1984). Immunochemical studies with the antipeptide antibody showed that the $S_2$-glycoprotein is localized extracellularly in the mucilage that surrounds the transmitting tissue cells and the wall of the ovule. This corresponds to the expression of the $S_2$-gene in transmitting tissue cells and the epidermis of the placenta demonstrated by in situ hybridization (Cornish et al., 1987). The secretion of the $S_2$-glycoprotein is consistent with the presence of signal sequences detected in two of the cDNA clones. The extra- cellular location in the mucilage of the transmitting cells and the ovule coincides with the pathway of pollen tube growth so that the pollen tubes would be in contact with the $S_2$-glycoprotein as they grow through the pistil.

The carbohydrate chains are another source of potential variation and hence of potential allelic specificity. Although the glycosylation sites are conserved, the numbers of chains present vary and the sequences within these chains may also vary. Detailed structural analysis of the individual chains is required to clarify this point.

Together, the DNA gel blot analyses of genomic DNA and the breeding data indicate that the three cDNAs described are indeed allelic products of the same gene. Each $S$-allele has a characteristic set of HindIII restriction fragments (Bernatzky et al., 1988) which hybridize most strongly to cDNA encoded by that allele and weakly with cDNA encoded by other $S$-alleles. This indicates that the $S_2^-$, $S_3^-$, and $S_6$-cDNAs are derived from the same locus. If they were from different loci, the DNA gel blot analyses would be expected to show bands corresponding to all genotypes in each track, unless there was specific deletion of genes corresponding to all but the expressed allele. Breeding experiments demonstrated that the allele-specific RFLPs segregated with $S$-genotype in plants produced by crossing two individuals with no alleles in common or with one allele in common, consistent with the interpretation that the $S_2^-$, $S_3^-$, and $S_6$-cDNAs represent alleles of the same locus or closely linked genes. Faint bands common to each genotype (see Figure 3A) were detected in DNA gel blot analyses; the significance of these bands cannot be interpreted until the corresponding DNA is cloned and sequenced. Comparison of the amino acid sequences of the three $N. alata$ $S$-glycoproteins with the amino acid sequences of $S$-glycoproteins from the sporophytically controlled self-incompatibility of *Brassica* spp. (Isogai et al., 1987; Nasrallah et al., 1987) shows that there is no homology between the sequences of the two groups of $S$-glycoproteins, suggesting that the two systems of self-incompatibility evolved separately. The allelic $S$-glycoproteins from *Brassica* species have a higher level of sequence homology than the three allelic $S$-glycoproteins from *N. alata*; like the $N. alata$ $S$-glycoproteins, the cysteine residues are conserved, but there is apparently more variation in the numbers of potential glycosylation sites.

In the absence of information on the corresponding allelic products for pollen, it is difficult to interpret the amino acid sequence data in functional terms. We are also cautious of drawing general conclusions from a relatively narrow data base; in some systems such as clover (Attwood, 1944; Williams, 1947), several hundred alleles exist. At present we have information on only a few of the many alleles of two self-incompatibility systems; as more se-
sequence data from related systems become available, the structure-function relationship will become clearer.

**METHODS**

**Plant Material**

Plants were maintained in a pollinator-proof glasshouse. S-Genotype was assigned by monitoring for seedset after hand pollination with pollen from tester plants homozygous for the \( S_a \), \( S_r \), \( S_s \), and \( S_p \)-alleles.

**cDNA Library Construction and Screening**

Poly-A+ RNA was isolated from mature \( S_a \) and \( S_p \)-alleles styles, and double-stranded cDNA was prepared using the Ribonuclease H method and Amersham Kit RPN1256. The cDNA was cloned into \( \lambda \text{gt}10 \) (Huynh, Young, and Davis, 1984), and \( S_a \) and \( S_p \)-clones were selected from the libraries by differential screening with single-stranded \(^{32}P\)-cDNAs prepared from style poly-A+ RNA from \( S_a \) and \( S_p \) genotypes. Prehybridization (2 hr at 68°C) was in 1.5 x SSPE, 1.0% SDS, 0.5% BLOTTO, and 0.5 mg/ml denatured herring sperm DNA (Bio-Rad Bulletin 1234). Hybridization was in prehybridization solution with added dextran sulfate (10% w/v) and \( S_a \)-cDNA library that hybridized with the \( S_a \)-cDNA but not the \( S_p \)-cDNA. One candidate clone, with the longest insert, was selected for each of the \( S_a \)-alleles. The \( S_a \)-cDNA library was screened in the same manner except that clones were selected which hybridized to \( S_a \)-cDNA clone NA-3-1 and were therefore the most likely candidates for cDNAs encoding the \( S_p \)-alleles. The \( S_p \)-cDNA library was screened in the same manner except that clones were selected which hybridized to \( S_p \)-cDNA and not \( S_a \)-cDNA. One candidate clone, with the longest insert, was selected for each of the \( S_p \) and \( S_s \)-alleles and used for subsequent work (NA-3-1, NA-6-1).

The cDNA inserts from clones NA-3-1 and NA-6-1 were ligated into either the bacteriophage vector M13mp18RF DNA for single-strand sequencing or into pGEM-3 (Promega) for double-strand sequencing using the chain termination method.

Sequence data obtained in this way were analyzed as described previously (Anderson et al., 1986).

DNA and RNA Gel Blots

DNA gel blot analysis was performed using 1.2% agarose-formaldehyde gels as described previously (Anderson et al., 1986). The DNA was transferred to hybrid bond-N (Amersham) and probed with either \(^{32}P\)-labeled \( S_a \)-cDNA (NA-2-2), \( S_p \)-cDNA (NA-3-1), or \( S_p \)-cDNA (NA-6-1) (3 x 10^6 cpm \( \mu \text{g}^{-1} \), 1 x 10^6 cpm \( \text{ml}^{-1} \)).

Prehybridization (2 hr, 68°C) was in 1.5 x SSPE, 1.0% SDS, 0.5% BLOTTO, and 0.5 mg/ml denatured herring sperm DNA (Bio-Rad Bulletin 1234). Hybridization was in prehybridization solution with added dextran sulfate (10% w/v final concentration).

Filters were washed twice in 2 x SSC, 0.1% SDS at 68°C, and once in 0.1 x SSC, 0.1% SDS at 68°C. The DNA was transferred to the membrane by "wet blotting" in 0.5 M NaOH. The membrane was then neutralized with 0.5 M Tris-HCl, 0.1% SDS, pH 7.5, followed by 2 x SSC, 0.1% SDS before hybridization at 68°C. Filters were washed in 1 x SSC, 0.1% SDS at 68°C and exposed to x-ray film at -70°C for 4 days. A HindIII digest of \( \lambda \) DNA was used as fragment-size markers.

**Peptide Synthesis, Antibody Production, and Purification**

The 22-amino acid peptide C-D-R-S-K-P-Y-N-M-F-T-D-G-K-K-K-N-D-L-D-E-R, starting from position 47 (Figure 2) was synthesized by the BOC-polystyrene solid-phase procedure using an Applied Biosystems Model 430A peptide synthesizer. The assembled peptide was cleaved from the resin with anhydrous hydrogen fluoride and the crude peptide purified by preparative reverse-phase HPLC using a \( \mu \)Bondapak C18 column (buffer A, 0.1% TFA-water, buffer B, 0.1% TFA-water-60% acetonitrile; 10% to 80% B over 60 min). The purified peptide was found to be homogeneous, as assessed by analytical HPLC and amino acid analysis. The yield from crude was 24%. The purified peptide (5 mg) in the reduced, monomeric form was conjugated to bovine thyroglobulin (0.5 mg) (Skowsky and Fisher, 1972). The conjugated peptide (1 mg/ml in PBS) was mixed with an equal volume of Freund's complete adjuvant. Three sheep were injected intramuscularly with 2 ml. After 4 weeks, booster injections containing 0.5 mg of peptide conjugated to 0.1 mg of thyroglobulin and emulsified into 1 ml of Freund's incomplete adjuvant were given to each of the three sheep. After 6 weeks, the sheep were bled. The antibody levels were assessed by radioimmunoassay using \(^{125}I\)-peptide as antigen (Skelly, Brown, and Besch, 1973).

The antibody was affinity purified on an \( S_p \)-peptide Sepharose column. The column was prepared by covalent attachment of 5 mg of \( S_p \)-peptide to 5 g of activated CH-Sepharose (Pharmacia LKB Biotechnology Inc.) according to the manufacturer's instructions. Sheep serum (5 ml) was applied to the column. Bound antibody was eluted using citric acid-phosphate buffer and was pooled and dialyzed against phosphate-buffered saline (yield, 7 ml, 400 \( \mu \text{g/ml} \)).

**Protein Extraction and Immunoblotting**

Style extracts were prepared by crushing three mature styles in 150 \( \mu \text{l} \) of extraction buffer (50 mM Tris-HCl, pH 8.5, 10 mM EDTA, 1 mM CaCl\(_2\), 1 mM phenylmethylsulfonfyl fluoride, 1 mM dithiothreitol). The extracts were microcentrifuged for 10 min at 4°C and 5 \( \mu \)l of the supernatant was removed and applied to an SDS-polyacrylamide gel (12.5%) in Laemmli sample buffer (Laemmli, 1970).

When electrophoresis was complete, the gel was electrophoblotted and the nitrocellulose was incubated with affinity-purified, anti-\( S_p \)-
peptide antibody (0.8 μg/ml in 1% gelatin in TBS [20 mM Tris-HCl, pH 7.5, 500 mM NaCl]). The nitrocellulose was washed with 0.05% Tween in TBS, and the bound antibody was visualized by incubation with rabbit anti-sheep IgG horseradish peroxidase conjugate (0.1 μg/ml in 1% gelatin in TBS, Bio-Rad), followed by washing in 0.05% Tween in TBS and incubation with horseradish peroxidase color development solution (Bio-Rad).

Immunocytochemistry

Styles (S₉S₉-genotype) were fixed as previously described (McFadden et al., 1988). For immunofluorescence, 1-μm sections were incubated with anti-S₉-peptide antibody (30 μg/ml) in SC buffer (McFadden et al., 1988) for 1 hr at 20°C. Sections were washed in SC, and then incubated with rabbit anti-sheep IgG conjugated to 15 nm colloidal gold (EY Laboratories). The control antibody, raised in sheep against the synthetic peptide arginine vasopressin and labeled with rabbit anti-sheep IgG antibody conjugated to 0.1 pg/ml in 1% gelatin in TBS (Bio-Rad), was a kind gift from Dr. Mario Congin.

ACKNOWLEDGMENTS

We thank Dr. M. Altschuler, Dr. R. Kalla, Dr. W. Jahnen, and Dr. R. Simpson for valuable discussions; Mr. B. McGinness and Ms. S. Mau for providing plant material; and Ms. M. Strode, Ms. A. Shuttleworth, and Ms. J. Noble for typing the manuscript.

Received March 13, 1989.

REFERENCES


**Sequence variability of three alleles of the self-incompatibility gene of Nicotiana alata.**  
M A Anderson, G I McFadden, R Bernatzky, A Atkinson, T Orpin, H Dedman, G Tregear, R Fernley and A E Clarke  
*Plant Cell* 1989;1:483-491  
DOI 10.1105/tpc.1.5.483

This information is current as of August 5, 2017

| --- | --- |
| eTOCs | Sign up for eTOCs at:  
http://www.plantcell.org/cgi/alerts/ctmain |
| CiteTrack Alerts | Sign up for CiteTrack Alerts at:  
http://www.plantcell.org/cgi/alerts/ctmain |
| Subscription Information | Subscription Information for *The Plant Cell* and *Plant Physiology* is available at:  
http://www.aspb.org/publications/subscriptions.cfm |