The Sequences of S-Glycoproteins Involved in Self-incompatibility of *Brassica campestris* and Their Distribution among Brassicaceae

Seiyei Yamakawa, Hiroshi Shibata, Masao Watanabe,* Hideyuki Shiozawa, Seiji Takayama, Kokichi Hinata,* Akira Isogai, and Akinori Suzuki

Department of Agricultural Chemistry, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan
* Faculty of Agriculture, Tohoku University, Aoba-ku, Sendai 981, Japan

Received December 1, 1993

S-Glycoproteins have been thought to play a role in the self-incompatibility in *Brassica* plants. Two cDNAs encoding S-glycoproteins were isolated from stigmas of two strains of *Brassica campestris*. Their amino acid sequences are highly homologous to each other. However, the numbers of the potential glycosylation sites are different, although most of them are conserved. In the middle part of the sequences, there are two supervariable regions which might contribute to the specificity of each S-glycoprotein. Genomic Southern analysis showed many bands hybridizing with cDNA of S-glycoproteins in various *Brassica* species including self-compatible species.

The self-incompatibility system in higher plants has created genetic diversity through the rejection of self-fertilization and the induction of outcrossing. In *Brassica*, the self-incompatibility system is under the sporophytic control of a single locus S with multi-alleles, and the rejection of self-pollination occurs at the surface of a stigma papilla cell when the same allele is active both in pollen and papilla cells.

Analysis of self-incompatibility in *Brassica* has been advanced by laying stress on the study of a class of stigma glycoproteins, the S-glycoproteins. Since the S-glycoproteins were detected immunologically and biochemically, several lines of evidence have been accumulated to suggest that the S-glycoproteins are involved in the expression of self-incompatibility. A cDNA encoding an S-glycoprotein of *Brassica* was first obtained from *B. oleracea* by Nasrallah et al. On the other hand, we isolated S5-, S9-, and S12-glycoproteins of *B. campestris* and partially analyzed the amino acid sequences. B. oleracea and *B. campestris* are closely related species and the S-glycoproteins of both species were quite similar. From these studies, the structural feature of the S-glycoproteins was found to be 53 kd secretory proteins with several saccharide chains. The structures of the saccharide chains have been established. In addition to these studies, several cDNAs have been sequenced from other strains of *B. oleracea* and the genomic DNA of *B. campestris* S8 strain has also been analyzed. Furthermore, the localization of these S-glycoproteins in the papilla cells was confirmed immunocytochemically. They were observed mainly in the cell walls of the papilla cells, where the pollen attached first. Thus, the S-glycoproteins have been thought to be key compounds to understand the mechanism of the self-incompatibility system.

During our studies on the structures of S-glycoproteins of *B. campestris*, we noticed the presence of NS(SRA)-glycoproteins. The SRA-glycoproteins were quite similar to the S-glycoproteins in structure and the pattern of localization in the papilla cells. But they cannot be classified as the S-glycoproteins because SRA-glycoproteins are independent of the S-genotype on crossing. In *B. campestris*, SRA-glycoproteins are 4 types controlled by allelic genes. Among them, we isolated and sequenced the cDNAs of SRA1- and SRA3-glycoproteins. Also in *B. oleracea*, quite similar glycoproteins were isolated by several laboratories. In this paper, we present the cDNA sequences encoding the S-glycoproteins that were isolated from two S-homozygotes, S9 and S12 of *B. campestris*, and discuss the homology among these S-glycoproteins. Furthermore, we examined the distribution of genes encoding the S- and S-related proteins in Brassicaceae by genomic Southern hybridization.

Materials and Methods

Plant materials. For construction of the cDNA library, two homozygotes, S9S9 and S12S12 of *B. campestris* were used. The stigmas were collected by the method previously reported, and the excised stigmas were stored at −80 °C until used. For genomic Southern analysis, *B. campestris* S8 homozygote, *B. oleracea* S1 homozygote, *B. nigra* 138, *Raphanus sativus* 28, *Diploitsis erucoides* 9, *Moricandia arvensis* 5, *Sinapis alba* 29, *Erucia sativa* 4, *B. napus* 127, *B. juncea* 473, and *B. carinata* 115 (numbers indicate the accession numbers in Tohoku University) were used. The latter three are self-compatible allo-tetraploid species. The mature leaves were excised, immediately frozen by liquid nitrogen, and stored at −80 °C until used.

Construction of the cDNA library. Total RNA from stigmas was isolated by centrifugation on GTC-CTC density gradients or by the SDS-phenol method. Poly(A)+RNA was isolated from total RNA using an oligo(T)-cellulose column.

SmaI cDNA was synthesized from poly(A)+RNA using cDNA Synthesis System Plus (Amersham) or ZAP-cDNA Synthesis Kit (Stratagene). The double-stranded cDNA was ligated to λ-gt10 or λ-ZAPII arms and cloned by in vitro packaging using a λDNA Cloning System λ-gt10 (Amersham) or GigapackII Packaging Extracts (Stratagene). As a result, the libraries involved approximately 2.2 × 107 recombinant clones.

Screening of the libraries and sequencing of the genes of S9- and S12-glycoproteins. To obtain the gene encoding S9 glycoprotein, the library was screened with two oligonucleotide probes, 31 and 62mers, synthesized.

---

* We have proposed a new system for naming the products of S- and S-related genes: SRA corresponds to NS or SLR1 which were previously reported; and SRA (B. campestris) is to be used for NS1.
on an Applied Biosystem 380B DNA synthesizer. These probes have antisense sequences designed on the basis of parts of the amino acid sequences of S₈-glycoprotein. Codon selection was based on the published cDNA sequence of SLG, which was isolated from an S⁸ homozygote of B. oleracea. Labeling of oligonucleotide probes and plaque screening on nylon membranes were done as described by Benton and Davis. Hybridization and washing conditions for each probe were as follows: 31mer: 6 x SSC, 5 x Denhardt's, 0.5% (w/v) SDS, 40°C for 15 h (hybridization); 2 x SSC, 0.1% SDS, 45°C for 90 min (washing). (20 x SSC is 0.3% NaCl and 0.3 M sodium citrate (pH 7.0). 10 x Denhardt's is 0.2% (w/v) PVP and 0.2% (w/v) BSA); 62mer: 50°C 15 h (hybridization); 55°C for 90 min (washing). Hybridization and washing solutions in the case of 62mer were the same as 31mers.

To screen the gene of S¹²-glycoprotein, the full length of the obtained S⁸-cDNA was used as a probe. Labeling of the probe was done by using a Random Primer DNA Labeling Kit (Takara) and screening was done by GeneScreen Plus (Du Pont). Hybridization and washing conditions were as follows: 1 x SSC, 1 M NaCl, and 10% dextran sulfate, 65°C for 15 h (hybridization); 1 x SSC, 1% SDS, 65°C for 40 min and 0.1 x SSC, 1% SDS, 65°C for 40 min (washing).

The selected cDNA clones were subcloned into the plasmid vector PAC18 or pBluescript and used to transform E. coli JM109 competent cells. The resulting recombinant plasmids were sequenced with Sequenase Ver. 2.0 (Toyobo) and a deletion kit (Takara) according to the supplier's instructions. The DNA sequences were established by reading both chains.

**Results and Discussion**

For screening of the cDNA encoding S⁸-glycoprotein, two probes were designed based on the partial amino acid sequence. The obtained cDNA library contained about 2.2 x 10⁶ independent clones. Eighteen thousand independent clones were screened and 8 clones hybridized with both of the probes. So they were expected to encode the S⁸-glycoprotein. Among them, the clone that showed the strongest signal was chosen for the sequence analysis. The size of the insert of the clone was estimated by agarose electrophoresis as about 1.4 kb, and it was confirmed that the clone had a sufficient size to encode the S⁸-glycoprotein. The cDNA clone was subcloned and fully sequenced. The length of this clone was 1506 bp. This clone encoded 436 amino acids starting with a translation initiation codon (ATG). Though this clone lacked the poly(A) tail at the 3' terminus, its deduced amino acid sequence contained all the fragments obtained from the S⁸-glycoprotein. And this nucleotide sequence included the full length of the genome sequence of S⁸-glycoprotein published by Nasrallah et al. So we concluded that this cDNA clone encoded the S⁸-glycoprotein. The characteristics of the sequence will be discussed later.

The cDNA encoding S¹²-glycoprotein was screened using the full-length of SLG²-cDNA as a probe. The size of the insert was confirmed by PCR and a clone that had adequate length was subcloned and sequenced as shown in Fig. 1. The length of this clone was 1472 bp. This clone encoded 436 amino acids starting with a translation initiation codon (ATG). The amino acid sequence deduced from the cDNA sequence contained all the peptide sequences including the N-terminal from the S¹²-glycoprotein. A signal sequence

**Fig. 1.** A cDNA Sequence Encoding S¹²-Glycoprotein and the Sequence of the Putative Amino Acid. The bars are under the amino acid sequence from the protein sequencing. The dashed line indicates the signal peptide and the putative polyadenylation signal.
S-Glycoproteins of *Brassica campestris* 923

(31 amino acids), following the ATG initiation codon, is present in the upstream region of the S^{12}-glycoprotein. The typical eucaryotic polyadenylation signals (AATAAA) and the poly(A) tail were found downstream of the termination codon (TAG).

The comparison of the amino acid sequences of S^{8} and S^{12}-glycoproteins are shown in Fig. 2A along with that of the S^{7}-glycoprotein, respectively. Thus the homology in the central region, Region 3, is lower than that in the other regions. The amino acid sequence of S-glycoproteins of *B. campestris* are quite homologous to those of *B. oleracea*\(^{10}\) and the variable regions of the S-glycoproteins of *B. oleracea* are the same as those of *B. campestris*. There are two supervariable sequence clusters in Region 3, from residues 182 to 189 and from residues 251 to 273. Hydrophilicity indices were almost identical among the three glycoproteins as expected from their homology. When the molecules of SLGs form the tertiary structures in the plant cells, perhaps the part around the Region 3 is exposed on the surface of the molecule and plays a key role in the structural differences among the various S-alleles. The two supervariable sequence clusters in the Region 3 are allelespecific, so these regions may play the most important roles of the SLGs as the product of the S-gene in self or non-self recognition. Among three glycoproteins, the S^{7}-glycoprotein has some differences from the other two. In the S^{7}-glycoprotein, amino acid deletions with more than two continuing residues were observed at three positions. And it has three additional amino acids at the C-terminus. This characteristic may show that the genetic distance between S^{9} and S^{12} is farther than that between S^{8} and S^{12}.

As for the potential glycosylation site, there are several sites (Asn-X-Ser or Asn-X-Thr sequence) distributed throughout the three S-glycoproteins as shown in Fig. 2B. Among them, six sites are common. These differences may be involved in the recognition system of S-glycoproteins. However, all the potential glycosylation sites do not actually bear the saccharide chains. In our preliminary data, most C-terminal sites of the S^{8}-glycoprotein did not have any saccharide chains.\(^{9}\) We will confirm the sites which actually bear the saccharide chains in S-glycoproteins using mass spectrometry.

The 12 cysteines at the C-terminal part were all conserved in three glycoproteins. These cysteines are well-known to be conserved along SLG or SLG-related members of the *Brassica* S-multigene family.\(^{9}\) These residues have been expected to execute an important role in the structure and/or function of SLGs.

In the comparison of the nucleotide sequence of SLG^{8}-cDNA with that of the genome sequence published by Nasrallah et al.,\(^{10}\) the cDNA sequence varied from the genome DNA in six nucleotides. These discrepancies are probably based on the difference of individual plants. There were four nucleotide substitutions and two nucleotide insertions or deletions. Among them, substitutions at three points were non-synonymous changes and they caused the amino acid conversion as shown in Fig. 2A, while the remaining changes of three points were synonymous or present in the non-coding region. In the comparison of three points of amino acid conversion with the amino acid sequence from the S^{8}-glycoprotein,\(^{6}\) two residues (Leu76, Leu168) coincide with that deduced from SLG^{8}-cDNA and one residue (Phe128) coincides with that of the genome sequence. These data suggest the possibility that plants which are classified as the same S-genotype have a variety of S-glycoproteins which show some trivial changes that have no effect on the S-phenotypes. As these difference were not in the supervariable regions we mentioned before, we expect that this change is unlikely to cause the change of
the S-genotype. So it is quite probable that plant strains that belong to the same S-genotype have a few differences in the DNA sequences from one another but not to the extent that they cause a change of the S-phenotype. Accumulation of these differences might be linked with the creation of new S-genotypes.

The result of genomic Southern analysis is shown in Fig. 3. In the case of using SLG<sup>8</sup> as a probe, many bands were observed in all samples including self-compatible species (Figs. 3A and B). Such a multiple band system was also observed in the genome of *B. oleracea*.<sup>27</sup> When the stringency was raised, the numbers of bands were decreased, but many bands still remained (Fig. 3B). This result shows that the many genes that have high homology to the SLG gene exist on genomes of *Brassica* and that SLG-like genes are widespread in Brassicaceae. On the other hand, when SRA<sup>1</sup> was used as a probe, the number of bands was much fewer than the case of SLG<sup>8</sup>, though signals were observed in most of the samples. In the comparison of the band patterns at high stringency between SLG<sup>8</sup> and SRA<sup>1</sup> (Fig. 3B and D), the bands of both lanes did not coincide with each other. This result shows that the series of SLG gene and SRA gene do not cross-react under the conditions of high stringency, so we can distinguish these two types of genes by this method.

As indicated above, there also exists a gene homologous to S-glycoproteins in the genomes of self-compatible *Brassica* species. Although it is not certain whether these genes are expressed in all self-compatible plants, some cDNAs encoding SLG-like proteins have been isolated from *Arabidopsis thaliana*.<sup>28</sup> which is a compatible plant. The possibility that SLGs are expressed also in self-compatible species presents us with a serious question on the role of SLG in the system of self-incompatibility. This may suggest that the SLG is just one of many factors involved in self-incompatibility or self-compatibility.

The multi-band pattern on genomic Southern analysis using the SLG<sup>8</sup>-cDNA indicates the presence of numerous genes homologous to SLG-gene in *Brassica*. In consideration of the fact that these genes did not hybridize with the SRA<sup>1</sup>-cDNA, these genes should be highly homologous to SLG and these genes might be divided into two types. One type is that which encodes SLG and SLG-like proteins and is expressed in mRNAs. The other genes are pseudogenes. Many researchers have recently pointed out that one of the SLG-like proteins is an S-receptor kinase (SRK).<sup>29,30</sup> SRK is a kind of receptor protein kinase and has a receptor domain resembling SLG. So SRK has been expected to take a part in signal transduction of self or non-self recognition on the occasion of pollination. The SRKs were shown to act in concert with the SLG on crossing.<sup>26</sup> Part of these multi-bands may show the SRK or another unknown
SLG-like gene, some of them being important for self-incompatibility. Whether all these genes including pseudogenes are in the S-locus or not are uncertain. For the sake of consideration of the roles of these pseudogenes in the process of diversification of the S-locus, it is significant to investigate the relation of segregation between S-alleles and these bands on crossing.

When SRA-cDNA was used as a probe in the genomic Southern analysis, the band patterns were simpler than the case of SLG. We have shown by immunoblot analysis that there are only four types of SRA-glycoproteins in *B. campestris*. This fact might be related to the simplicity of band patterns.

The functions of S-glycoproteins are uncertain. SLG-genes will be just one of the products of the S-gene and there will be more products on the S-locus that will carry out the important roles in self-incompatibility. Until now genomic analysis is partly reported for a strain of *B. oleracea*. So it is necessary to analyze the genome of *B. campestris* to understand the complete perspective of the S-locus of the *Brassica* species.

Acknowledgments. This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan, and Takeda Science Foundation.

References