Identification of regions in which positive selection may operate in S-RNase of Rosaceae: Implication for *S*-allele-specific recognition sites in S-RNase

Takeshi Ishimizu[a], Toshinori Endo[b], Yumi Yamaguchi-Kabata[c], Kazuo T. Nakamura[d], Fumio Sakiyama[a], Shigemi Norioka[a,*]

[a]*Division of Protein Chemistry, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan*
[b]*Center for Information Biology, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan*
[c]*Department of Biophysics, Faculty of Science, Kyoto University, Kyoto 606-8502, Japan*
[d]*School of Pharmaceutical Sciences, Showa University, Hatanodai, Shinagawa, Tokyo 142-8555, Japan*

Received 14 September 1998; received in revised form 28 October 1998

Abstract **A stylar S-RNase is associated with gametophytic self-incompatibility in the Rosaceae, Solanaceae, and Scrophulariaceae. This S-RNase is responsible for *S*-allele-specific recognition in the self-incompatible reaction, but how it functions in specific discrimination is not clear. Window analysis of the numbers of synonymous (d_S) and non-synonymous (d_N) substitutions in rosaceous S-RNases detected four regions with an excess of d_N over d_S in which positive selection may operate (PS regions). The topology of the secondary structure of the S-RNases predicted by the PHD method is very similar to that of fungal RNase Rh whose tertiary structure is known. When the sequences of S-RNases are aligned with the sequence of RNase Rh based on the predicted secondary structures, the four PS regions correspond to two surface sites on the tertiary structure of RNase Rh. These findings suggest that in S-RNases the PS regions also form two sites and are candidates for the recognition sites for *S*-allele-specific discrimination.**

Key words: Self-incompatibility; S-RNase; Rosacea; Positive selection; Non-synonymous nucleotide substitution

1. Introduction

Self-incompatibility is a mechanism that prevents self-fertilization in flowering plants. Rosaceous species such as *Pyrus pyrifolia* (Japanese pear) and *Malus x domestica* (apple) have gametophytic self-incompatibility controlled by a single, multi-allelic locus, the *S*-locus [1,2]. When a pollen grain lands on a stigma, a discrimination process takes place as to whether the *S*-allele of the pollen matches one of the two *S*-alleles of the pistil. The pollen grain germinates on the stigma and grows toward the embryo; but, if its *S*-allele matches one of the *S*-alleles of the pistil, pollen tube growth is arrested in the style, and no fertilization take places. A pistil-specific protein encoded by the *S*-locus has been shown to be a ribonuclease (S-RNase) that recognizes the pollen *S*-allele [3–7]. S-RNase-based gametophytic self-incompatibility also operates in the Solanaceae and Scrophulariaceae [8,9]. Two models for the *S*-allele-specific inhibition of pollen tube growth involving S-RNase, the *S*-allele-specific uptake and RNase inhibitor models, have been proposed by researchers working on self-incompatibility [8,9], but at present there is not enough evidence to support either model. How S-RNase discriminates between self- and non-self-pollen, what the counterpart molecule interacting with S-RNase is, and how S-RNase interacts with that molecule have yet to be clarified.

Rare *S*-alleles have a reproductive advantage because pollen bearing such alleles is less likely to land on a stigma with the same allele, and many kinds of *S*-alleles are maintained in a finite population [10]. Overdominant selection (heterozygote advantage) therefore is considered to occur at the *S*-locus in the population of the genus. Recently, recognition sites in some proteins (e.g. major histocompatibility complex (MHC) [11–13], antigenic surface proteins of parasites and viruses [13], and acrosomal proteins of abalone [14,15]) have been reported to be regions in which the number of non-synonymous nucleotide substitutions (d_N) exceeds that of synonymous substitutions (d_S), and positive selection probably takes place in these regions. Proteins related to self-incompatibility, including S-RNase, will have such a region if overdominant selection operates in them [16].

The primary structural features of solanaceous S-RNases have been studied. Pairwise comparisons show that these S-RNases have highly divergent amino acid sequences and that some interspecific pairs have higher sequence similarities than the intraspecific pairs [17]. The level of constraint on nucleotide substitution is heterogeneous throughout the S-RNase gene, some regions being highly constrained and others virtually unconstrained. Window analysis of d_N and d_S, however, failed to detect a region in which d_N exceeds d_S [18]. Moreover, no region with an excess of d_N over d_S was detected in S-related Brassicaceae proteins [19]. No recognition sites have yet been identified even in series of transgenic experiments using chimeric genes between two S-RNases [9,20].

Several cDNAs encoding S-RNases from rosaceous species (*P. pyrifolia* and *M. x domestica*) in Rosaceae have been cloned [21–26]. These S-RNases are distinct from the solanaceous and scrophulariaceous S-RNases and other RNase T_2 family enzymes in sequence alignment [21,23,24] and in the neighbor-joining phylogenetic tree [27]. To identify the *S*-allele-specific recognition site in S-RNase, we conducted window analysis of d_S and d_N in 11 rosaceous S-RNases and found four regions with an excess of d_N over d_S in them. Moreover, their secondary structures predicted by the PHD method were very similar to the RNase Rh structure, the only

*Corresponding author. Fax: (81) (6) 879-8619.
E-mail. norioka@protein.osaka-u.ac.jp

Abbreviations: d_S, the number of synonymous substitutions; d_N, the number of non-synonymous substitutions; HV region, hypervariable region; PS region, putative positively selected region

PII: S0014-5793(98)01470-7

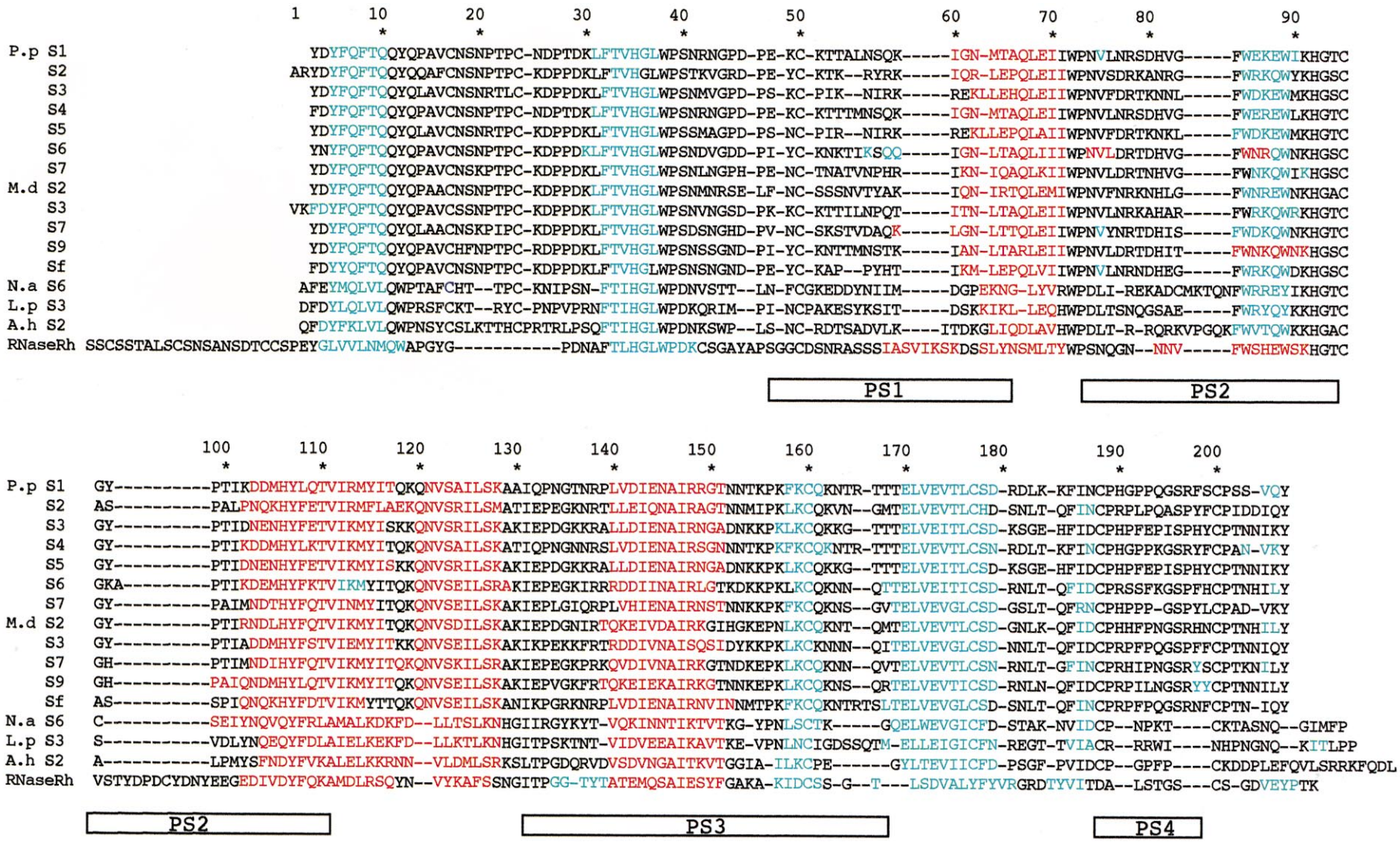

Fig. 1. Predicted secondary structures of rosaceous S-RNases (P.p, *Pyrus pyrifolia*; M.d, *Malus x domestica*). These structures were predicted by the PHD method [30], as were the structures of the solanaceous and scrophulariaceous S-RNases, *Nicotiana alata* (N.a) S_6- [29], *Lycopersicon peruvianum* (L.p) S_3- [28], and *Antirrhinum hispanicum* (A.h) S_2-RNases [27]. Residues predicted to form the α-helix are shown in red and those that form the β-strand in blue. The amino acid sequence of RNase Rh was aligned manually with the S-RNase sequences. The secondary structure of RNase Rh shown here was derived from its tertiary structure [33]. Four PS regions in the rosaceous S-RNases (described later) are indicated by boxes under the alignment.

known tertiary structure ribonuclease in the RNase T_2 family. We here discuss *S*-allele-specific recognition sites in the rosaceous S-RNases on the basis of the above findings.

2. Materials and methods

2.1. Sequences of S-RNases

The nucleotide and amino acid sequences of the rosaceous S-RNases used were *P. pyrifolia* S_2- and S_4-RNases (D49527 and D49528) [23]; S_1-, S_3-, S_5-, S_6-, and S_7-RNases (AB002139, AB002140, AB002141, AB002142, and AB002143) [26]; *M. x domestica* S_2-, S_3-RNases (U12199, U12200) [21], S_7-, S_9-RNases (U19792, U19793) [22]; and S_f-RNase (D50837) [24]. Those of the solanaceous S-RNases used are *Lycopersicon peruvianum* S_3-RNases (X76065) [28]; *Nicotiana alata* S_6-RNase (U08861) [29]; and *Antirrhinum hispanicum* S_2-RNases (X96465) [27].

2.2. Prediction of secondary structures

Secondary structures of the rosaceous, solanaceous, and scrophulariaceous S-RNases were predicted by the PHD method [30].

2.3. Sequence alignment and window analysis of d_S and d_N

On the basis of the predicted secondary structures, the amino acid sequences of the S-RNases were aligned manually with the sequence of *Rhizopus niveus* RNase Rh (D12476) [31]. The d_S and d_N values for each window of 20 codons along the aligned sequences in the S-RNase pairs were estimated [13,32]. The averages of the values for each window for all 55 pairs of the 11 S-RNases were plotted against the location of the window. Site number 1 on the horizontal column in Fig. 2 shows the window from codon number 1 to 20.

3. Results

3.1. Prediction of the secondary structures of S-RNases

Secondary structures of 12 rosaceous, two solanaceous, and one scrophulariaceous S-RNases were predicted by the PHD method [30] (Fig. 1). Positions of the predicted α-helical and β-stranded regions were similar. Many conserved residues appear to participate in the formation of these structures. The topology of the predicted secondary structures coincided with the structure of RNase Rh, the only known tertiary structure ribonuclease in the RNase T_2 family [33]. The amino acid sequences of the S-RNases were aligned with the sequence of RNase Rh using the predicted secondary structures (Fig. 1). The frameworks of the S-RNases also are likely to be very similar to the framework of RNase Rh because the secondary structure forms its core structure.

3.2. Window analysis of d_S and d_N

The averages of d_S and d_N for each window of 20 codons for all 55 pairs of the seven *P. pyrifolia* and four *M. x domestica* S-RNases are plotted against the location of the window (Fig. 2). Sequence alignment is the same as in Fig. 1. Four statistically significant regions in which d_N was higher than d_S were detected in site (amino acid) numbers 38–55 (48–65), 63–100 (73–110), 121–158 (131–168), and 179–188 (189–198). These regions respectively were designated PS1, PS2, PS3, and PS4 (abbreviation of putative positively selected region). A markedly higher d_N than in the other regions was found in the PS1 region. This region includes the HV region (amino acid numbers 51–66) which has many amino acid substitutions in rosaceous S-RNases and is thought to be one of the *S*-allele-specific recognition sites [26]. The HVa and HVb regions in solanaceous S-RNases [34] respectively correspond to the PS1 and PS2 regions described above (Fig. 1). The four PS regions are shaded in Fig. 1.

3.3. Location of the PS regions in S-RNase

On the assumption that the tertiary structure of S-RNase is similar to that of RNase Rh, the PS1, PS2, PS3, and PS4 regions correspond to two surface sites on RNase Rh (Fig. 3a) based on the sequence alignment in Fig. 1. The PS1 and PS2 regions form one site on the left side of the tertiary structure of RNase Rh in Fig. 3a. The PS3 and PS4 regions

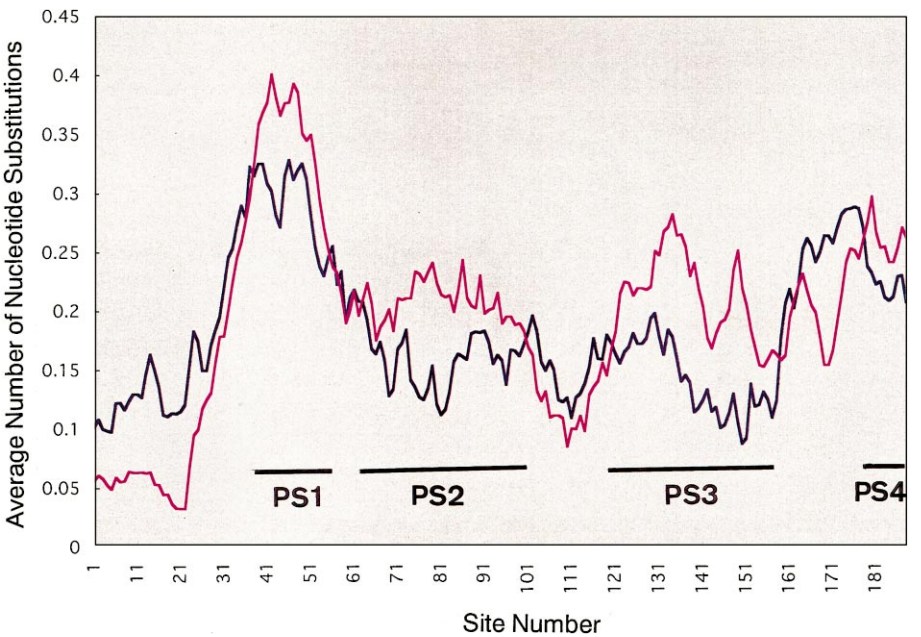

Fig. 2. Window analysis of synonymous and non-synonymous nucleotide substitutions in maloideous S-RNases. The window is defined as a sequence region 20 codons long. Averages of the synonymous and non-synonymous nucleotide substitutions for all 55 pairs of seven *Pyrus pyrifolia* and four *Malus x domestica* S-RNases were plotted against the location of the window. Synonymous substitutions are indicated by the blue line, non-synonymous ones by the red one.

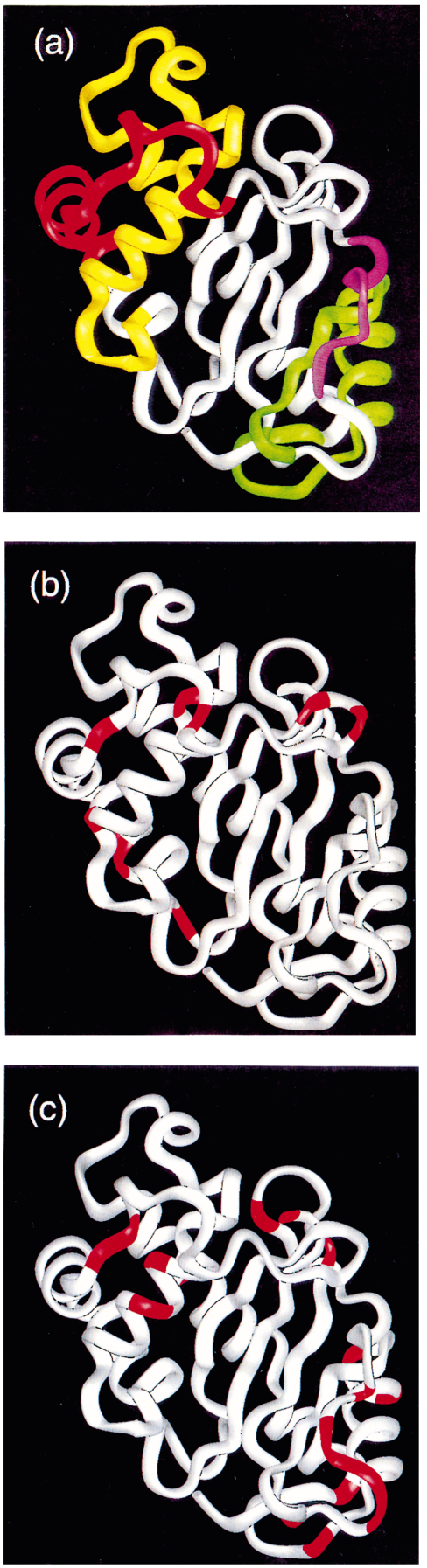

Fig. 3. Locations of the four PS regions of rosaceous S-RNases and the amino acid substitutions between the two highly homologous pairs of S-RNases on the tertiary structure of RNase Rh. The PS1, PS2, PS3, and PS4 regions (a) respectively are colored red, yellow, green, and purple. Amino acid substitutions between (b) *Pyrus pyrifolia* S_3- and S_5-RNases and (c) *Pyrus pyrifolia* S_1- and S_4-RNases are shown in red.
←

form the other site on the right side. The substrate RNA-binding cleft is located between these two sites.

Positions of the amino acid substitutions in two highly homologous pairs of S-RNases are marked on the tertiary structure of RNase Rh (Fig. 3b,c). All nine amino acid substitutions between the *P. pyrifolia* S_3- and S_5-RNase pairs (96% sequence identity) are located in the PS1 and PS2 regions or their adjacent substrate-binding cleft (Fig. 3b). In the pair of *P. pyrifolia* S_1- and S_4-RNases (90% sequence identity), 11 of the 20 substitutions are in or around the PS3 and PS4 regions (Fig. 3c), six being located in or around the PS1 and PS2 regions.

4. Discussion

S-RNase is associated with the stylar component of gametophytic self-incompatibility. It cosegregates with the *S*-locus, and is believed to function in the recognition of *S*-allele type. To identify the *S*-allele-specific recognition sites and to explore how S-RNase acts in the self-incompatible reaction, we analyzed d_S and d_N in 20-codon windows for all pairs of 11 rosaceous S-RNases. Four regions in which d_N exceeded d_S were detected in rosaceous S-RNases (Fig. 2). This is the first report of regions with an excess of d_N over d_S in self-incompatible-related proteins. No such region was detected in solanaceous S-RNases [18]. This difference may be due to the ages of the respective gene groups because the d_S values of solanaceous S-RNases are about two-fold those of rosaceous S-RNases [18]. In solanaceous S-RNases, however, the d_N values of regions corresponding to the four PS regions described above are slightly higher than those of the other regions [18]. In addition, no region with an excess of d_N over d_S has been detected in proteins associated with sporophytic self-incompatibility, although some regions had high d_N values [19].

The d_N value in the PS1 region was considerably higher than in the other regions, even including the other three PS regions. Insertion, deletion, or recombination of genes as well as nucleotide substitution perhaps occur frequently in this region. Interestingly, the only intron in the S-RNases is inserted in this region, between the 59th and 60th amino acid residues (N. Norioka, unpublished results). The higher d_N values in the PS1 compared to other PS regions may reflect different functions. For example, several base-pair differences in the HV region may be required for *S*-allele-specific discrimination.

Regions with an excess of d_N over d_S are present in the antigen recognition sites of MHC proteins [11–13], antigenic surface proteins of parasites and viruses [13], and acrosomal proteins of abalone [14,15]. The four regions in S-RNase also probably function as recognition sites that interact with certain pollen molecules, and positive selection may take place in these regions.

The similarity between the secondary structures determined for RNase Rh and predicted for the S-RNases suggests similarity at the tertiary level as well. The tertiary structures of the rosaceous S-RNases were predicted by the 3D-1D method, using the program COMPASS [35,36], which indicated a strong similarity to RNase Rh (K. Nishikawa and M. Ota, personal communication). The PS1 and the PS2 regions together correspond to a surface site in the tertiary structure of RNase Rh (Fig. 3a), and the PS3 and PS4 regions form the opposite surface site. The question arises whether the two sites of S-RNase interact with a single molecule or whether each site interacts with separate molecules. Among the molecules that may interact with S-RNase are proteins encoded at the *S*-locus that are expressed specifically in pollen or products of modifier genes distinct from the *S*-locus [37,38], although no such products have been identified as yet.

The PS regions have many polar, especially basic, amino acid residues. That arginine and lysine residues tend to be located on the surface of the protein [39] is consistent with the PS regions being located on the surface of S-RNase. These basic amino acid residues may function in the interaction with a counterpart molecule.

The sequence identity between the *P. pyrifolia* S_3- and S_5-RNases is very high (96%), and only nine amino acid substitutions are located in or around the site formed by the PS1 and PS2 regions (Fig. 3b). This site therefore may be responsible for discriminating between S_3 and S_5 pollens and for triggering the self-incompatible reaction. Recently, it has been reported that four amino acid substitutions in the PS1 and PS2 regions between *Solanum chacoense* S_{11}- and S_{13}-RNases are necessary and sufficient to discriminate between S_{11} and S_{13} pollen [40]. In this pair, one amino acid substitution (from Met to Val) also exists in the PS3 and PS4 regions, but such a small difference may be insufficient to discriminate pollen *S*-allele type. The *P. pyrifolia* S_1- and S_4-RNases also are highly homologous (90% sequence identity). Although the 20 amino acid substitutions between this pair are dispersed throughout the primary structure, their positions are located in or around the two sites on the tertiary structure of RNase Rh (Fig. 3c). These findings support the hypothesis that the HV and PS regions function as recognition sites for the discrimination of *S*-alleles and do not contradict the report that *S*-allele specificity is not determined by a single segment of S-RNase [20].

Our findings do not conflict with the classical genetic theory of gametophytic self-incompatibility [10]. A newly functional *S*-allele introduced into a population is preferred to other *S*-alleles because pollen bearing the new *S*-allele is less likely to match one of the two *S*-alleles of the pistil. Loss by random genetic drift of a new *S*-allele occurs less often than loss of a neutral allele, and the frequency of the new *S*-allele is increased until equilibrium is reached. Such new *S*-alleles must be generated by non-synonymous substitutions at *S*-allele-specific recognition sites on the *S*-locus. The PS regions, which have high d_N values, therefore are speculated to be *S*-allele-specific recognition sites. The functions of the PS regions must now be investigated by biochemical and biological methods to clarify the molecular mechanism of S-RNase-based gametophytic self-incompatibility.

Acknowledgements: We are grateful to Ken Nishikawa and Motonori Ota (National Institute of Genetics, Japan) for predicting the tertiary structures of the S-RNases by the 3D-1D method and valuable suggestions. This work was supported in part by a Grant-in-Aid for Scientific Research on a Priority Area (Genetic Dissection of Sexual Differentiation and Pollination Process in Higher Plants, No. 07281103) from the Ministry of Education, Science, Sports and Culture of Japan and by Research Fellowships for Young Scientists provided by the Japan Society for the Promotion of Science.

References

[1] Kikuchi, A. (1929) J. Okitsu Hortic. Soc. 24, 1–6 (in Japanese).
[2] Latimer, L.P. (1937) Proc. Am. Soc. Hortic. Sci. 34, 19–21.
[3] Anderson, M.A., Cornish, E.C., Mau, S.-L., Williams, E.G., Hoggart, R., Atkinson, A., Bonig, I., Grego, B., Simpson, R., Roche, P.J., Haley, J.D., Penschow, J.D., Niall, H.D., Tregear, G.W., Coghlan, J.P., Crawford, R.J. and Clarke, A.E. (1986) Nature 321, 38–44.
[4] McClure, B.A., Haring, V., Ebert, P.R., Anderson, M.A., Simpson, R.J., Sakiyama, F. and Clarke, A.E. (1989) Nature 342, 955–957.
[5] Lee, H.S., Huang, S. and Kao, T.-h. (1994) Nature 367, 560–563.
[6] Murfett, J., Atherton, T.L., Beiquan, M., Gasser, C.S. and McClure, B.A. (1994) Nature 367, 563–566.
[7] McCubbin, A.G., Chung, Y.-Y. and Kao, T.-h. (1997) Plant Cell 9, 85–95.
[8] Dodds, P.N., Clarke, A.E. and Newbigin, E. (1996) Cell 85, 141–144.
[9] Kao, T.-h. and McCubbin, A.G. (1996) Proc. Natl. Acad. Sci. USA 93, 12059–12065.
[10] Wright, S. (1939) Genetics 24, 538–552.
[11] Hughes, A.L. and Nei, M. (1988) Nature 335, 167–170.
[12] Hughes, A.L. and Nei, M. (1989) Proc. Natl. Acad. Sci. USA 86, 958–962.
[13] Endo, T., Ikeo, K. and Gojobori, T. (1996) Mol. Biol. Evol. 13, 685–690.
[14] Swanson, W.J. and Vacquier, V.D. (1995) Proc. Natl. Acad. Sci. USA 92, 4957–4961.
[15] Lee, Y.-H., Ota, T. and Vacquier, V.D. (1995) Mol. Biol. Evol. 12, 231–238.
[16] Maruyama, T. and Nei, M. (1981) Genetics 98, 441–459.
[17] Ioerger, T.R., Clarke, A.G. and Kao, T.-h. (1990) Proc. Natl. Acd. Sci. USA 87, 9732–9735.
[18] Clark, A.G. and Kao, T.-h. (1991) Proc. Natl. Acad. Sci. USA 88, 9823–9827.
[19] Hinata, K., Watanabe, M., Yamakawa, S., Satta, Y. and Isogai, A. (1995) Genetics 140, 1099–1104.
[20] Zurek, D.M., Mou, B., Beecher, B. and McClure, B. (1997) Plant J. 11, 797–808.
[21] Broothaerts, W., Jansserns, G.A., Proost, P. and Broekaert, W.F. (1995) Plant Mol. Biol. 27, 499–511.
[22] Janssens, G.A., Goderis, I.J., Broekaert, W.F. and Broothaerts, W. (1995) Theor. Appl. Genet. 91, 691–698.
[23] Norioka, N., Ohnishi, Y., Norioka, S., Ishimizu, T., Nakanishi, T. and Sakiyama, F. (1995) Plant Physiol. 108, 1343.
[24] Sassa, H., Nishio, T., Kowyama, Y., Hirano, H., Koba, T. and Ikehashi, H. (1996) Mol. Gen. Genet. 250, 547–557.
[25] Sassa, H. and Hirano, H. (1997) Plant Physiol. 113, 306.
[26] Ishimizu, T., Shinkawa, T., Sakiyama, F. and Norioka, S. (1998) Plant Mol. Biol. 37, 931–941.
[27] Xue, Y., Carpenter, R., Dickinson, H.G. and Coen, E.S. (1996) Plant Cell 8, 805–814.
[28] Royo, J., Kowyama, Y. and Clarke, A.E. (1994) Plant Physiol. 105, 751–752.
[29] Anderson, M.A., McFadden, G.I., Bernatzky, R., Atkinson, A., Orpin, T., Dedman, H., Tregear, G., Fernley, R. and Clarke, A.E. (1989) Plant Cell 1, 483–491.
[30] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[31] Horiuchi, H., Yanai, K., Takagi, M., Yano, K., Wakabayashi, E., Sanda, A., Mine, S., Ohgi, K. and Irie, M. (1988) J. Biochem. 103, 408–418.
[32] Nei, M. and Gojobori, T. (1986) Mol. Biol. Evol. 3, 428.
[33] Kurihara, H., Nonaka, T., Mitsui, Y., Ohgi, K., Irie, M. and Nakamura, K.T. (1996) J. Mol. Biol. 255, 310–320.

[34] Ioerger, T.R., Gohlke, J.R., Xu, B. and Kao, T.-h. (1991) Sex. Plant Reprod. 4, 81–87.
[35] Matsuo, Y., Yamada, A., Tsukamoto, K., Tamura, H., Ikezawa, H., Nakamura, H. and Nishikawa, K. (1996) Protein Sci. 5, 2459–2467.
[36] Ota, M. and Nishikawa, K. (1997) Protein Eng. 10, 339–351.
[37] Uyenoyama, M.K. (1991) Genetics 128, 452–469.
[38] Ai, Y., Kron, E. and Kao, T.-h. (1991) Mol. Gen. Genet. 230, 353–358.
[39] Miller, S., Janin, J., Lesk, A.M. and Chothia, C. (1987) J. Mol. Biol. 196, 641–656.
[40] Matton, D.P., Maes, O., Laublin, G., Xike, Q., Bertand, C., Morse, D. and Cappadocia, M. (1997) Plant Cell 9, 1757–1766.