Computational and Structural Biotechnology Journal 21 (2023) 5228–5239

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Comprehensive computational analysis of the SRK–SP11 molecular interaction underlying self-incompatibility in Brassicaceae using improved structure prediction for cysteine-rich proteins

Tomoki Sawa [a], Yoshitaka Moriwaki [a,b,*], Hanting Jiang [a], Kohji Murase [c], Seiji Takayama [c], Kentaro Shimizu [a,b], Tohru Terada [a,b]

[a] Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
[b] Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
[c] Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan

ARTICLE INFO

Keywords:
Self-incompatibility
Structure prediction
Cysteine-rich proteins
Multiple sequence alignment

ABSTRACT

Plants employ self-incompatibility (SI) to promote cross-fertilization. In Brassicaceae, this process is regulated by the formation of a complex between the pistil determinant *S* receptor kinase (SRK) and the pollen determinant *S*-locus protein 11 (SP11, also known as *S*-locus cysteine-rich protein, SCR). In our previous study, we used the crystal structures of two eSRK–SP11 complexes in *Brassica rapa* S_8 and S_9 haplotypes and nine computationally predicted complex models to demonstrate that only the SRK ectodomain (eSRK) and SP11 pairs derived from the same *S* haplotype exhibit high binding free energy. However, predicting the eSRK–SP11 complex structures for the other 100 + *S* haplotypes and genera remains difficult because of SP11 polymorphism in sequence and structure. Although protein structure prediction using AlphaFold2 exhibits considerably high accuracy for most protein monomers and complexes, 46% of the predicted SP11 structures that we tested showed < 75 mean per-residue confidence score (pLDDT). Here, we demonstrate that the use of curated multiple sequence alignment (MSA) for cysteine-rich proteins significantly improved model accuracy for SP11 and eSRK–SP11 complexes. Additionally, we calculated the binding free energies of the predicted eSRK–SP11 complexes using molecular dynamics (MD) simulations and observed that some *Arabidopsis* haplotypes formed a binding mode that was critically different from that of *B. rapa* S_8 and S_9. Thus, our computational results provide insights into the haplotype-specific eSRK–SP11 binding modes in Brassicaceae at the residue level. The predicted models are freely available at Zenodo, https://doi.org/10.5281/zenodo.8047768.

1. Introduction

Self-incompatibility (SI) is a mechanism that many flowering plants acquire to prevent self-fertilization and promote genetic diversity [1,2]. Pollen hydration, tube germination, and elongation are rejected if the pollen is genetically identical or closely related to the pistils. Over the course of evolution, plants have acquired various molecular mechanisms for SI that are currently being researched in the Brassicaceae [3–5], Solanaceae [6–9], and Papaveraceae [10–13] families among others. In Brassicaceae, SI is sporophytically controlled by the *S*-locus [3–5]. The family Brassicaceae contains diverse genera and species for which extensive surveys of the diversity of the *S*-locus have been performed, including *Brassica rapa* (Chinese cabbage, turnip, etc.), *B. oleracea* (cabbage, broccoli, etc.), *B. napus* (rapeseed, etc.), *Raphanus sativus* (radish, etc.), *R. raphanistrum* (wild radish, etc.), *Arabidopsis lyrata*, and *A. halleri*.

S-locus glycoprotein (SLG) was the first identified product of the *S*-locus [14,15]. Subsequently, a receptor-type kinase called *S*-receptor kinase (SRK) with an SLG-like extracellular region was identified as the pistil determinant of SI in *B. oleracea* (*Bo*) and *B. rapa* (*Br*) [16,17]. Furthermore, *S*-locus protein 11 (SP11 or *S*-locus cysteine-rich protein (SCR)) has been identified as a determinant present on the pollen side [18–21]. SRK, which is located on the plasma membrane of the papilla cells, specifically binds to its cognate SP11 that is released from the

* Corresponding author at: Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 1138657, Japan.
E-mail address: moriwaki@bi.a.u-tokyo.ac.jp (Y. Moriwaki).

https://doi.org/10.1016/j.csbj.2023.10.026
Received 11 July 2023; Received in revised form 3 October 2023; Accepted 16 October 2023
Available online 20 October 2023

pollen surface. This triggers the self-phosphorylation of the kinase domain inside the cell, which results in the rejection of fertilization [22–26]. The alleles of *SLG*, *SRK*, and *SP11* are tightly linked at the *S*-locus and are transmitted to the progeny as a single set called *S* haplotype [18,27].

The amino acid sequences of SRK and SP11 are significantly different among the *S* haplotypes. The SRK ectodomain (eSRK) shares two lectin domains with an EGF-like domain and a PAN domain among the haplotypes [28] in addition to three hypervariable (HV I–III) regions with notably different sequence compositions [29,30]. In contrast, SP11 is a defensin-like protein that comprises 60–70 amino acids with four or five disulfide bonds and is characterized by numerous insertions, deletions, and variations across haplotypes [31]. Based on sequence similarity, the *S* haplotypes are classified into classes I and II in *Brassica* and *Raphanus*. *Br* class II comprises the S_{29}, S_{40}, S_{44}, and S_{60} haplotypes [32]. However, the dominant nature of class I pollen phenotype hinders that of class II. Regulation of pollen dominance is achieved through the suppression of SP11 expression by two small RNAs: *SP11 methylation inducer* (*SMI*) and *SMI2* [33,34].

Recently, two eSRK–SP11 complex crystal structures have been reported. Ma et al. determined the *Br* S_9-eSRK–S_9-SP11 complex structure, in which the two SP11 molecules are tightly bound to the interface of the two eSRKs [35]. Subsequently, we presented the crystal structure of the engineered *Br* S_8-eSRK–S_8-SP11 complex and nine computational complex models of S_8-, S_9-relative and class-II haplotypes using homology modeling and accelerated molecular dynamics (MD) simulations [36]. Additionally, we showed that the calculated binding free energies were largely negative for all 11 cognate eSRK–SP11 complexes but not for the 156 non-cognate ones. We have also validated that a point mutation in the S_{36}-SP11 residue located at the interface, inferred from the S_{36}-eSRK–S_{36}-SP11 complex model, disrupted the SI reaction in the pollination bioassay [36]. These results indicate that the use of appropriate predicted structures can lead to a better understanding of the self/nonself-discrimination mechanism between eSRK and SP11 without experimental crystallization. However, attaining a comprehensive understanding of the mechanism encompassing all *S* haplotypes remains challenging because the estimated number of haplotypes is > 100 for *Br*[4] and > 50 for *Bo* [37]. Additionally, most SRK and SP11 proteins cannot be readily expressed in vitro, making experimental validation highly difficult. Furthermore, obtaining the stable structures for all eSRK–SP11 pairs through accelerated MD simulations is impractical owing to the substantial computational resources required.

AlphaFold2 [38] and AlphaFold-Multimer (AF-Multimer) [39] released in 2021, exhibited high accuracy in predicting both monomeric and complex structures. We present a concise overview of how AlphaFold2/AF-Multimer performs structure prediction: First, multiple sequence alignment (MSA) for the input sequence is retrieved from the sequence database using jackHMMER [40] and HHblits [41], and template structures are searched in the Protein Data Bank (PDB) database using HHSearch [42]; second, the Evoformer processes the MSA and template structures to extract their converged evolutionary and spatial relationships (denoted as MSA and pair representations in that paper); third, the Structure Module assembles the protein structure based on the representations, and the output structure is reutilized thrice as the input of the Evoformer. Importantly, the accuracy of the predicted model decreases substantially when the median MSA depth is < 30 sequences. This underscores the substantial influence of both the quantity and quality of MSA on their accuracy.

The great success of AlphaFold2 in the field of protein structure prediction is expected to advance our understanding of proteins based on their structures. ColabFold [43], which is a derivative of AlphaFold2, offers structure prediction through a web browser by leveraging the MMSeqs2 webserver [44,45] to obtain an MSA file of the input sequence. Moreover, starting from July 2022, AlphaFold Protein Structure Database (AlphaFold DB) [46] has made predicted monomeric structures available for nearly all proteins recorded in the UniProt database. Despite the per-residue model confidence (a pLDDT metric) [38] being reliably high for eSRK model structures in AlphaFold DB, the confidence for SP11 models is relatively low. This discrepancy is attributed to the limited number of homologous sequences available for SP11. An additional crucial fact worth mentioning is that the AlphaFold DB does not provide structural information for protein complexes that are crucial for understanding molecular recognition.

In this study, we report more plausible SP11 and eSRK–SP11 model structures using the cognate eSRK and SP11 sequence pairs that were annotated in previous experimental studies to overcome the drawbacks of structure prediction with AlphaFold2. Moreover, we comprehensively explore the molecular recognition mechanism between cognate eSRK and SP11 pairs at the residue level based on predicted complex models. Consequently, the models suggested that the presence of variable regions other than HV I–III, which may contribute to the self/nonself-discrimination, and the eSRK–SP11 binding mode vary significantly depending on the *S* haplotypes.

2. Methods

2.1. Preparation of amino acid sequences

The amino acid sequences were obtained from GenBank for eSRK and SP11 from seven species, namely, *Brassica rapa* (*Br*), *B. oleracea* (*Bo*), *B. napus* (*Bn*), *R. sativus* (*Rs*), *R. raphanistrum* (*Rr*), *A. lyrata* (*Al*), and *A. halleri* (*Ah*) and used for structural prediction. In total, 98 eSRK and SP11 sequence pairs derived from the same *S*-haplotypes were used in this study [17–19,21,30–32,47–70]. The sequences and accession IDs are shown in **Supplementary File 1**.

2.2. Structure prediction with ColabFold

The structures of eSRK and SP11 proteins were predicted using ColabFold [43], which is a derivative of AlphaFold2. During a typical prediction process, ColabFold uses the MMSeqs2 webserver to retrieve an MSA file of the input amino acid sequence [44,45]. In this study, we obtained MSA files using the UniRef30 [71] (uniref30_2202) and BFD/MGnify (bfd_mgy_colabfold) databases. We denoted the retrieved MSA files as "default" MSA for the comparison described in a later section. In the prediction using ColabFold, the recycle number for the prediction was set to three and the template structure was not used. These predictions were performed using LocalColabFold, which is the command line interface of ColabFold (https://github.com/YoshitakaMo/localcolabfold).

2.3. Construction of custom MSAs for ColabFold

In addition to the "default" MSA, four other MSAs were constructed to improve the predicted models. The CysBar [72] Python script was used to construct MSAs that properly expressed the structural features of the SP11 defensin-like domain. This tool facilitates the alignment of structurally homologous cysteine residues within the MSA. Typically, such homologous Cys residues are identified by combining them with known structures of high homology detected using the DALI server [73,74]. However, owing to the large number of structurally unknown and polymorphic SP11s, the disulfide bond pattern of the S_8- or S_9-SP11 crystal structures was used as the standard, and other SP11s were assumed to form similar pairs. Of the 98 SP11 sequences, the type with eight cysteine residues was the most prevalent; thus, these eight residues were designated as disulfide bonds at positions 1–8, 2–5, 3–6, and 4–7. For *Br* S_{46}-SP11 with seven Cys residues and S_{32}-/S_{36}-SP11 with ten Cys residues, the pairs were inferred from a previous study [36]. The other SP11s whose sequences closely resembled them were assumed to comprise the same pairs. After the application of CysBar, the 98 sequences were processed using Clustal Omega [75,76] to obtain "cys-seq" MSA. For comparison, "seq" MSA was similarly obtained without using

CysBar. The resultant "seq" and "cys-seq" MSAs are shown in Fig. S1.

Additionally, two structure-based MSAs, namely, "st" and "st2" MSAs, were constructed using the 98 sequences for comparison with the sequence-based alignments. The St MSA was constructed as follows:

1. The two crystal structures for SP11 S_8 and S_9 and 16 modeled SP11 structures for *Br* S_{12}, S_{21}, S_{25}, S_{29}, S_{32}, S_{34}, S_{36}, S_{40}, S_{44}, S_{45}, S_{46}, S_{47}, S_{49}, S_{52}, S_{60}, and S_{61}, some of which were reported in our previous study [36], were subjected to Mustang 3.2.3 [77] to obtain their structure-based MSA.
2. Each of the remaining 80 sequences were added to the MSA by creating pairwise alignments with one of the 18 sequences that they were most closely related to according to the phylogenetic tree (Fig. S2).

St2 MSA was created as follows:

1. First, all 98 SP11 structures were predicted using ColabFold with the "default" MSA.
2. Second, the structures were aligned using Mustang [77].

Each of the four MSA files comprised 98 sequences. We ran the structure prediction using the MSA file with LocalColabFold by moving the corresponding sequence to the top of the file and reformatting it into the a3m format using the "reformat.pl" script, which is bundled in the HH-suite [42]. No templates were used for predictions using the four MSAs.

The normalized number of effective sequences (N_{eff}) was computed for each position of a query sequence at a threshold of 80% sequence identity to estimate the quality of the MSA, which is also described as MSA depth in AlphaFold2 [38].

2.4. Improvement of the structure prediction for class-II SP11s

For class-II SP11s, a more confident model was predicted using ColabFold with a distinct MSA composed of class-II SP11 sequences retrieved from the BLASTP search as the input and a predicted structure that showed the best pLDDT among the four MSAs as the template. The recycling number for the prediction was increased to 9.

2.5. Evaluation of predicted models

We used the pLDDT confidence measure [38], which is a predicted per-residue lDDT-Cα score [78], to evaluate the predicted monomeric model of eSRK or SP11. Additionally, we used the metrics—pTM-score [38] and pDockQ [79]—to evaluate the predicted model for the eSRK–SP11 complex. The pLDDT and pTM-scores were calculated automatically using ColabFold for each predicted model, and we ranked the models according to the mean pLDDT for all residues per model. The pDockQ score was calculated using "pDockQ2.py" available at https://gitlab.com/ElofssonLab/huintaf2/-/blob/main/bin/pDockQ2.py and averaged for the chains in the model.

2.6. Paired MSA for the prediction of eSRK–SP11 complex

To run the structure prediction for a complex with proteins A and B, ColabFold requires an a3m-formatted MSA file in which a horizontally concatenated MSA of protein pairs A and B that bind correctly is placed in the first block, followed by the MSAs of proteins A and B. In this study, we prepared an MSA for each of the 98 Clustal Omega-created eSRK sequences [75,76] and for the four types of SP11 sequences described in the previous section. We concatenated the aligned eSRK and SP11 sequences for each *S* haplotype in the "paired" block of the a3m file. Aligned homologous sequences of eSRK and SP11 with gaps were placed in the second and third blocks, respectively. The format of the input a3m file is shown in Fig. S3.

2.7. MM–GBSA analysis

Molecular mechanics with generalized Born and surface area solvation (MM–GBSA) [80] implemented in AmberTools 22 [81] was used to calculate the binding free energy, ΔG_{bind}, between the modeled eSRK–SP11 complexes. The initial coordinates were obtained from a predicted complex model that exhibited the best pDockQ values for each haplotype. The ff19SB force field [82] was used for the proteins, and the system was solvated using the OPC water model [83]. Energy minimization, equilibration, and a 50-ns conventional production run were performed using AMBER 22 [81] as per the procedure described in our previous paper [36]. The GBOBC implicit solvent model (parameters $\alpha = 1.0$, $\beta = 0.8$, and $\gamma = 4.85$) [84] was used with a salt concentration of 0.2 M. All calculations were performed using MD trajectories between 10 and 50 ns and recorded every 100 ps (400 snapshots for each complex).

3. Results

3.1. eSRK and SP11 structures predicted using "default" MSA

We first used the MSAs retrieved from ColabFold/MMSeqs2 webserver, denoted as the "default" MSA, to predict the structures of eSRK and SP11 (Fig. 1**A**). Of the 98 haplotypes, 97 predicted eSRK structures and 53 predicted SP11 structures exhibited an average pLDDT of ≥ 75. The high pLDDT of eSRKs could be attributed to their long amino acid sequence length (approximately 400) and high sequence homology between the haplotypes, whereas the low pLDDT of SP11 could have been caused by the short amino acid length and many diverse insertions and deletions in the defensin-like SP11 domain in the MSA, making the prediction difficult. The relatively low pLDDT observed in *Rs* S_{22}-, *Rs* S_{26}-, and *Rs* S_{30}-SRK were due to the incomplete N- or C-terminal sequences (Fig. S4). Fig. 1**B** shows the correlation between N_{eff} and the average pLDDT of SP11 structures predicted using the default MSA. The average pLDDT increased as N_{eff} approached 100, which is consistent with the discussion of AlphaFold2 [38]. However, five outlier proteins, namely, *Br* S_{12}-, *Br* S_{33}-, *Rs* S_{23}-, *Rs* S_{31}-, and *Bo* S_{11}-SP11, exhibited low pLDDT despite having $N_{\mathrm{eff}} > 500$ because most of the MSAs are composed of inappropriate amino acid sequences that are partially analogous but not homologous to the full-length SP11. In contrast, only a limited number of homologous sequences could be obtained from the webserver for *Br* S_{36}- and *Bo* S_{24}-SP11, which was insufficient to obtain a successful prediction. The left panels of Fig. S5A and **B** visualize the MSA coverage of the default MSA for *Br* S_{12}- and *Br* S_{36}-SP11, respectively.

3.2. Custom MSAs improved SP11 models

ColabFold enables other MSA files to be specified as input instead of the default MSA. To obtain improved SP11 model structures, we prepared four types of MSA and denoted them as "cys-seq", "seq", "st", and "st2"; each consisted of 98 SP11 sequences that were experimentally annotated (see Methods). All four custom MSAs yielded SP11 models with a pLDDT score equivalent to or higher than that of the default MSA (Fig. 1**C and D**). For 66 of the 98 SP11s, the prediction accuracy was improved when one of the four MSAs was used. The median pLDDT increased to 84.0 by choosing the best of the four MSA models, and the number of SP11s showing pLDDT >75 increased from 53 to 81. Particularly, the *Br* S_{12}-SP11 model, which was predicted using cys-seq MSA, showed the largest improvement (46.0–89.5 in the mean pLDDT; Fig. 1**E**). Similarly, the prediction with custom MSAs also improved for *Br* S_{33}, *Bo* S_{11}, *Rs* S_{23}, and *Rs* S_{31} (Fig. S6).

Although the four MSAs improved the pLDDT scores for many SP11s, they had little effect for *Rs* S_{11} (48.5), S_{15} (52.3), and S_{26} (47.3) because their C-terminal sequences after the fifth Cys residue corresponding to the β3 strand are missing (Fig. S1). Similarly, the pLDDT scores of *Br* S_{46}, S_{61}, *Bo* S_7, S_{13}, and S_{13b} were <70 despite their sequences being close to

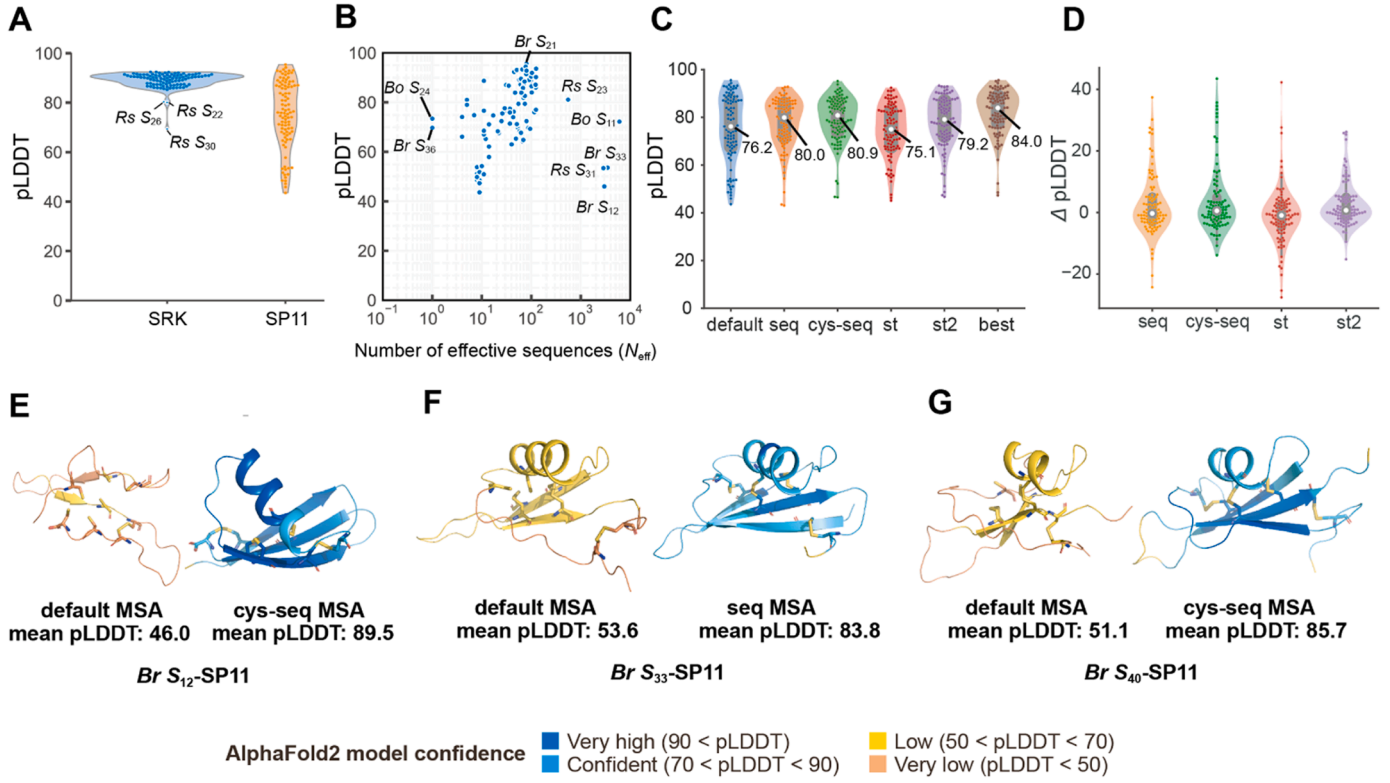


Fig. 1. Structure prediction of S-locus protein 11 (SP11) proteins with ColabFold using various multiple sequence alignments (MSAs). (A) The mean predicted local distance difference test (pLDDT) of *S* receptor kinase (SRK) and SP11 models predicted using ColabFold with the "default_MSA" (uniref30_2202 + bfd_mgy_colabfold database). (B) The mean pLDDT of SP11 models for the MSA depth. N_{eff} was calculated according to a previously reported procedure [94] with a threshold of 62% sequence identity. (C) The mean pLDDT of the predicted SP11 models vs. the four custom MSAs ($n = 98$). The white circle represents the median. (D) Difference in pLDDT with respect to the default MSA. The white circle represents the median. (E–G) Significantly improved models of *Br* S_{12} (E), *Br* S_{33} (F), and *Br* S_{40} (G). Models predicted with the default and one of the four MSAs which exhibits the best pLDDT are shown in left and right panels, respectively.

Br S_8, the crystal structure of which was determined. All pLDDT scores for the predicted models obtained using the default and four MSAs are provided in **Supplementary File 2**.

3.3. Structural novelty in the predicted SP11 proteins

Our predicted models with high pLDDT values enable us to explore

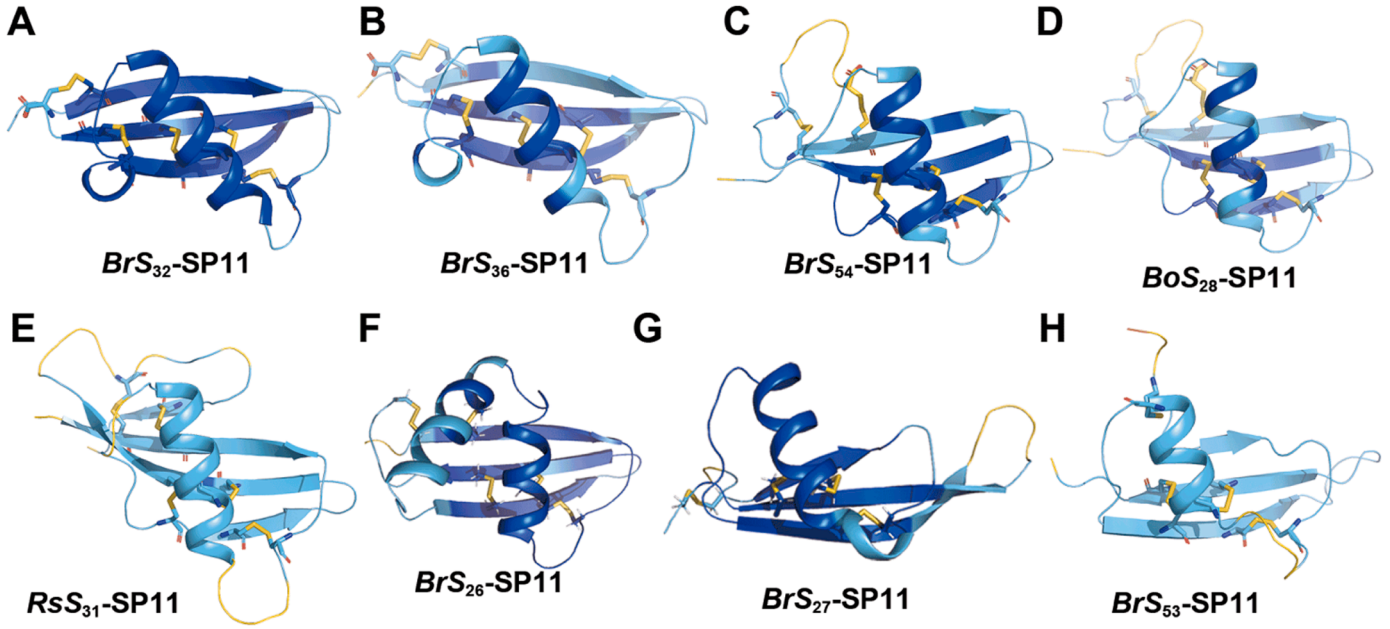


Fig. 2. Predicted novel class I SP11 structures and disulfide bond connectivity. (A–H) *Br* S_{32} (A), *Br* S_{36} (B), *Br* S_{54} (C), *Bo* S_{28} (D), *Rs* S_{31} (E), *Br* S_{26} (F), *Br* S_{27} (G), *Br* S_{53} (H)-SP11 model structures predicted using ColabFold with custom MSAs. Disulfide bonds are shown as sticks. The models are colored according to the AlphaFold2 model confidence.

the polymorphisms of SP11 proteins based on their sequences and structures. Although most SP11s typically possess eight Cys residues and the disulfide pairs are conserved to be 1–8, 2–5, 3–6, and 4–7 [85], we found three structural groups with different disulfide bond pairs in SP11s with 10 Cys residues. First, *Br* S_{32} and *Br* S_{36} showed 1–10, 2–6, 3–7, 4–9, and 5–8 disulfide bond connectivity (Fig. 2**A–B**), in which two disulfide bonds exist between the β2 and β3 strands. The bond connectivity and regions of the secondary structures of *Br* S_{32} and *Br* S_{36} were accurately predicted in our previous study using homology modeling and accelerated MD simulations [36] (Fig. S7). The predicted SP11 models of *Bo* S_{24} and *Bo* S_{68} also exhibited the same connectivity, and they closely resembled *Br* S_{36} and *Br* S_{32}, respectively, because their sequence similarities were substantially high (Fig. S2). Second, *Br* S_{54}, *Bo* S_{28}, and *Rs* S_{31} exhibited 1–9, 2–10, 3–6, 4–7, and 5–8 disulfide bond connectivity, in which two Cys residues on the β1 strand form bonds with those located after the β3 strand (Fig. 2**C–E**). Notably, SP11 of *Rs* S_{31} formed a β4 strand at its C-terminus (Fig. 2**E**). Third, *Br* S_{26}-SP11 showed the 1–10, 2–5, 3–7, 4–8, and 6–9 connectivity (Fig. 2**F**). It possesses two helices between the β1 and β2 strands, although the second helix is not involved in the interaction with its corresponding eSRK (Fig. S8).

Furthermore, two anomalous structures are worth mentioning in the SP11s with eight Cys residues: the first is *Br* S_{27} and its one-residue variant, *Bo* S_8; both exhibited long twisted β-hairpin structures (Fig. 2**G**). The second is *Br* S_{53}, which exhibited a structure comprising 1–4, 2–6, 3–7, and 5–8 disulfide connectivity (Fig. 2**H**); the fourth Cys residue was located at the end of the α helix.

Of the class-II 10 *S* haplotypes, *Br* S_{29}, *Br* S_{40}, *Br* S_{44}, *Br* S_{60}, *Bo* S_5, *Rr* S_6, and *Rs* S_9 exhibited a mean per-residue pLDDT of ≥ 75 (Fig. 3**A–G, middle**), and their backbone structures resembled each other (Fig. 3**H**). The prediction for the remaining three haplotypes, *Rs* S_{11}, *Rs* S_{15}, and *Rs* S_{26}, still failed for the reason described in the previous section. Applying our refinement protocol for class-II SP11s (see Methods) to the seven haplotypes with accurately predicted structures improved their pLDDT values to > 87 (Fig. 3**A–G, right**). The predicted class-II structures possessed a characteristic β-bridge secondary structure after the first β-strand (Fig. 3**I and J**). The α-helix of *Br* S_{60} was longer than that of these haplotypes by two residues (Fig. 3**D**). Remarkably, only the *Rs* S_9 model lacked the α-helix between the β1 and β2 strands among the 98 SP11s (Fig. 3**G**) regardless of its high mean pLDDT score (91.7). All seven predicted structures showed substantially higher pLDDT values than of those presented in the AlphaFold DB version 2022–11–01 (Fig. 3**A–G, left**), indicating that new predicted structures were more plausible. The low confidence of the models in the AlphaFold DB could be attributed to the failure to detect the class-I SP11 sequences, which are evolutionarily distant from those of class-II sequences and to the inappropriate alignment of cysteine-rich protein fragments during the construction of the MSA.

3.4. Improved eSRK–SP11 complex prediction using ColabFold and paired MSAs

We also investigated the effect of the four MSAs on the modeling of the eSRK–SP11 complexes. We used three criteria to assess the plausibility of the predicted eSRK–SP11 complex models: the pTM-score [38], pDockQ score [79], and a root-mean-square-deviation (rmsd) for the crystal structure of the *Br* S_8 eSRK–SP11 complex (PDB:6KYW). Of the 98 predicted complex models with their default MSAs, only 41 showed pDockQ > 0.6. However, the prediction with a paired MSA, where one of the four SP11-MSAs are concatenated with the eSRK-MSA (see Fig. S3), led to 80 models meeting the criteria. (Fig. 4**A**). Notably, st2 and seq MSAs significantly enhanced the pDockQ scores of *Br* S_{34} and S_{55} complexes, respectively. Consequently, their two SP11 molecules were located between the eSRK dimers, similar to the conformation observed in the *Br* S_8/S_9 crystal structure (Fig. 4**C and D**). Similarly, *Br* S_{12} and *Br* S_{47} complex models were also improved by cys-seq and st MSAs, respectively (Fig. 4**E and F**). Overall, the four paired MSAs led to an enhanced pDockQ score for 95 complex models (Table 1), and the most effective MSAs varied depending on the *S* haplotype (see **Supplementary File 2**).

Of the 80 improved complex models, 76 exhibited rmsd < 5 Å, indicating that they share the same binding mode as that of *Br* S_8 or *Br* S_9. The complex predictions for *Rs* S_{11}, *Rs* S_{15}, *Rs* S_{26} were not successful owing to their poor SP11 models as described earlier. Remarkably, among the prediction for *Ah* and *Al*, *Ah* S_{28}, *Al* S_6, *Al* S_{14}, *Al* S_{18}, and *Al* S_a (also known as *Al* S_{13}) showed rmsd > 10 Å despite pDockQ > 0.50, although their monomeric structures were predicted with high pLDDT values (Fig. 4**B**). They formed a novel binding interface in which an SP11 molecule was placed between HV-II of one eSRK and HV-III of the other eSRK (Fig. 5**A**). In contrast, *Ah* S_{12}, *Ah* S_{20}, *Ah* S_{32}, *Al* S_{36}, and *Al* S_b (also known as *Al* S_{20}) showed a complex mode similar to that of the *Br* S_8 or *Br* S_9 crystal structures, in which HV-II was in direct contact with HV-III (Fig. 5**B**). The predicted aligned error (PAE) indicates the unreliability of the residue–residue relationship of the predicted AlphaFold2 model [38]. PAE was generally low for both models (Fig. 5**C and D**), suggesting that these predicted models were both plausible.

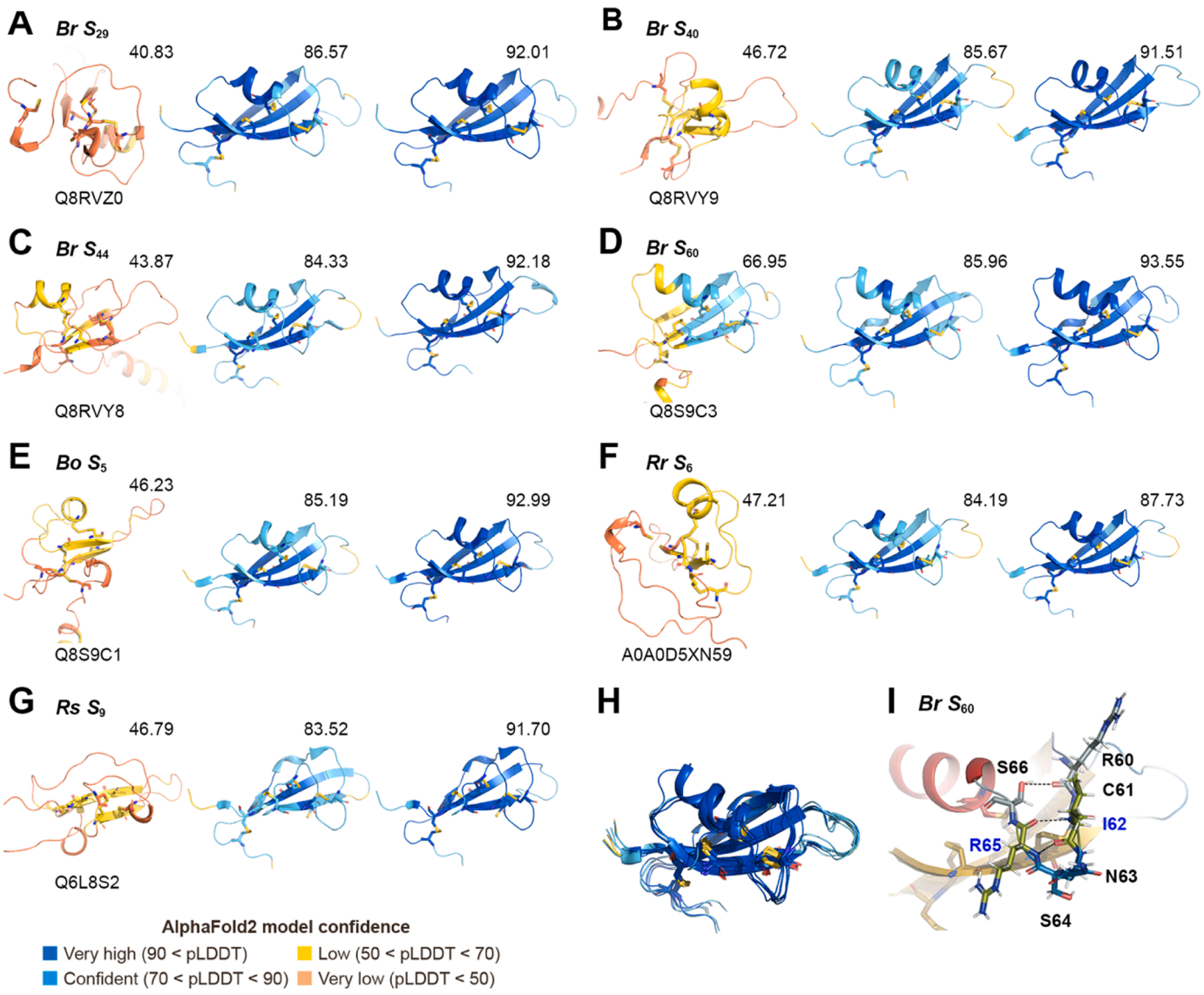
3.5. Binding free energy calculations for the predicted eSRK–SP11 complexes using MM–GBSA analysis

In this study, we obtained highly plausible eSRK–SP11 complex models for most haplotypes using ColabFold and curated MSAs. However, they do not guarantee tight interactions because the AlphaFold2/ColabFold prediction is statistical and not physical. Hence, we used MM–GBSA analysis [80] to calculate the binding free energy (ΔG_{bind}) for the predicted models.

Of the 98 complex models that were examined, 57 structures demonstrated $\Delta G_{bind} < -75$ kcal/mol. Fig. 6**A** illustrates the per-residue contribution to ΔG_{bind} on a sequence alignment of class-I *Br* SP11s arranged using CysBar [72] and Clustal Omega [75,76]. In general, the residues comprising the first loop region, the α helix, and the β2-β3 hairpin exhibited large negative ΔG_{bind}, which conforms with the results of our previous study [36] and the binding interfaces observed in the predicted complex models. The MM–GBSA analysis showed largely negative ΔG_{bind} values for haplotypes whose SP11 showed a characteristic secondary structure; in *Br* S_{26}, the contribution of ΔG_{bind} to eSRK interactions was observed near the kink between the first and second helices (Fig. S8); in *Br* S_{27}, another SP11 molecule interacts with its unique β2–β3 structure. These results demonstrate that most of the predicted complex models are truly bound with high affinity. The per-residue energy contributions for all haplotypes that showed a large negative ΔG_{bind} in this study are shown in **Supplementary File 3**.

Based on the predicted complex structures and MM–GBSA analysis, we propose the presence of two additional variable regions in eSRK: the N- and C-terminal accessorial variable regions. These regions are partly involved in the self/nonself-discrimination mechanism. For example, in *Br* S_{48}, Tyr69 and Phe70 that are located in the N-terminal accessorial variable region interacted with the α-helix of SP11, and Glu401 and Glu403 that are located in the C-terminal region formed salt bridges with Arg21 and Arg23 of SP11, respectively (Fig. 6**B**). Importantly, these regions do not necessarily contribute to the eSRK–SP11 interactions depending on the haplotype, as was not observed in the crystal structures of *Br* S_8 and S_9 (see also **Supplementary File 3**). Particularly, the C-terminal variable region was previously discovered by Kusaba et al. but it was not believed to be involved in the specificity between SRK and SP11 [29]. However, the predicted models and analyses suggest that this contributes to complex stabilization in some *S* haplotypes.

The predicted class-II complex structures presented a different binding mode from that observed for class-I. The amino acid compositions of β3 in class-II SP11s are well conserved, and a conserved Tyr residue at position 56 of SP11 interacted with the insertion characteristic of HV I in class-II eSRK, except in *Rr* S_6 (Fig. 6**C–E and** Fig. 3**H**).



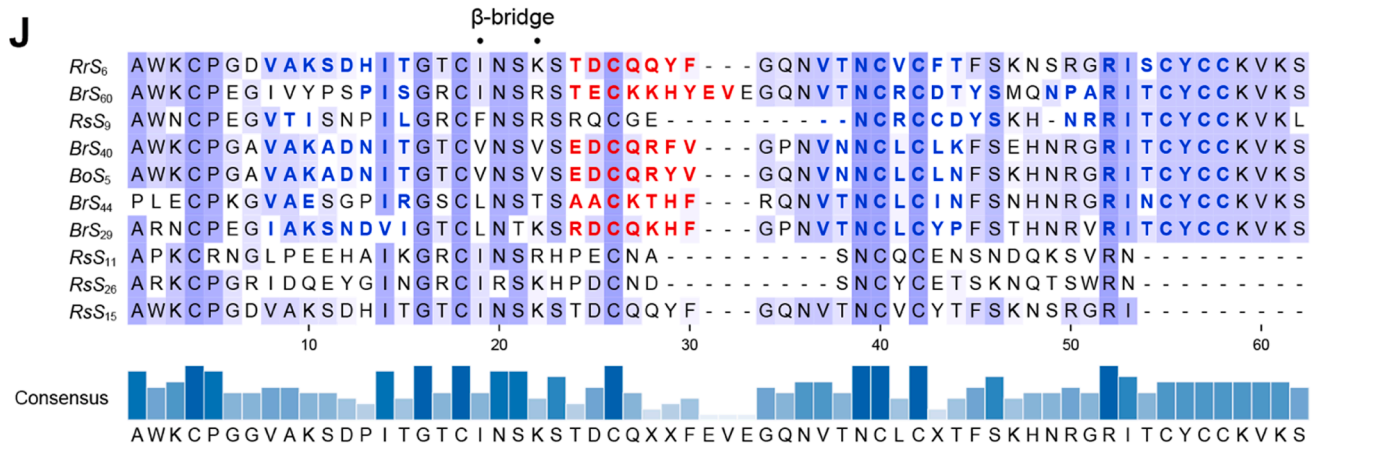


Fig. 3. Predicted models of class-II SP11s. (A–G) Predicted SP11 models of *Br* S_{29} (A), *Br* S_{40} (B), *Br* S_{44} (C), *Br* S_{60} (D), *Bo* S_5 (E), *Rr* S_6 (F), and *Rs* S_9 (G) haplotypes. The model deposited in AlphaFold DB (version 2022–11–01), the best one among the four MSAs, and one improved using the refinement protocol are depicted in the left, middle, and right panels, respectively. UniProt accession IDs are shown under the model of AlphaFold DB. The structures are colored according to the AlphaFold pLDDT model confidence. The mean per-residue pLDDT is shown above each model on the upper right side. (H) Superposition of the seven predicted class-II SP11 proteins. (I) A characteristic β-bridge structure observed in the predicted class II SP11 proteins. *Br* S_{60}-SP11 is shown as an example. The hydrogen bonds are shown as yellow dashed lines. (J) Sequence alignments of class-II SP11 proteins. The α-helix and β-strand secondary structures annotated using DSSP 4.0.5 are shown as red and blue characters, respectively. The characteristic β-bridge are indicated using "·" marks.

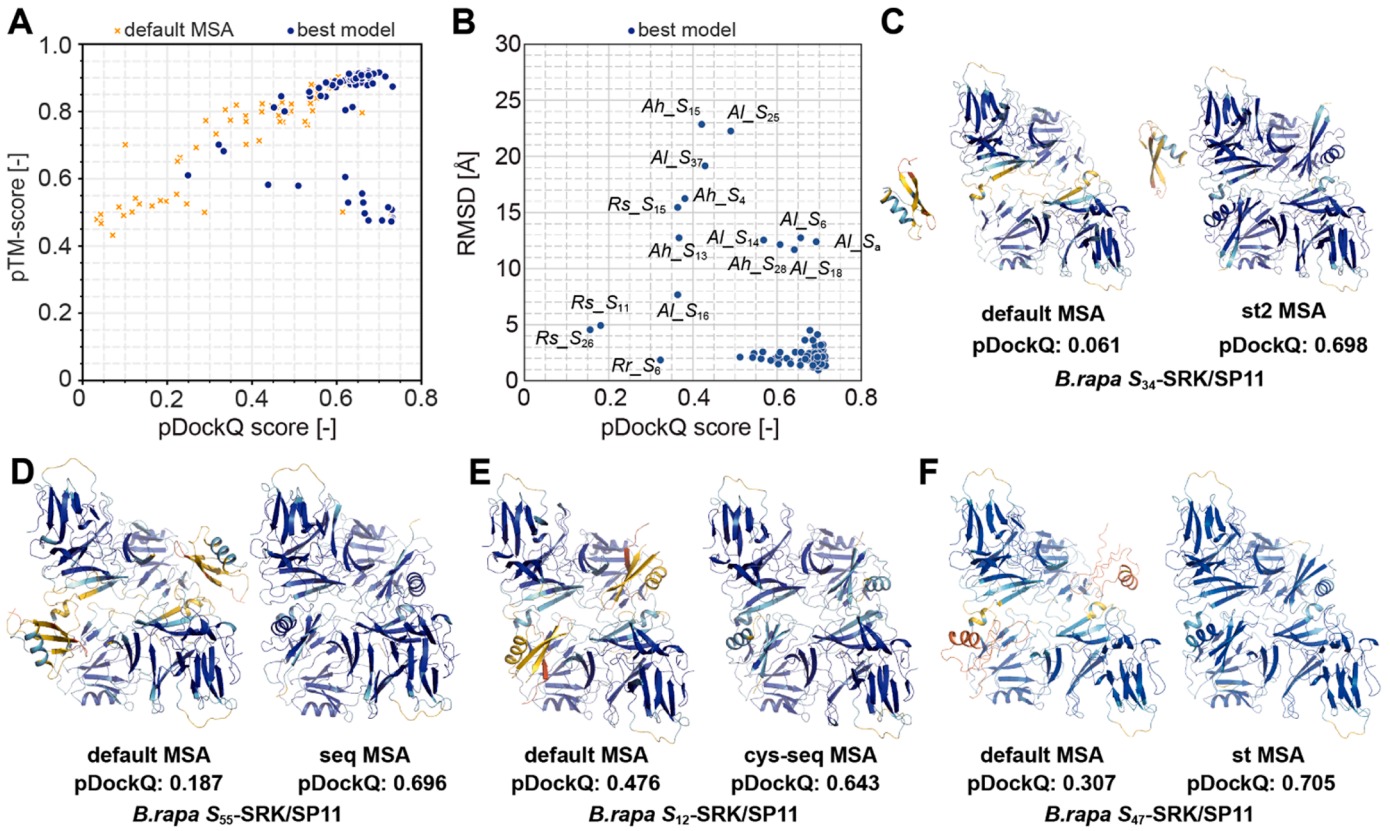


Fig. 4. Complex predictions with the custom MSAs. (A) Correlation between pTM-score and pDockQ for the predicted complex models. Orange crosses and blue circles indicate the complex models predicted using the default MSA and the highest pDockQ of the four MSAs, respectively. (B) Root mean square deviation (RMSD) values of the predicted model showing the highest pDockQ (best model) to the *Br* S_8 crystal structure (PDB: 6KYW). (C–F) Improved complex models of *Br* S_{34} (C), *Br* S_{55} (D), *Br* S_{12} (E), and *Br* S_{47} (F) haplotypes using the custom MSAs. The residues are colored based on the pLDDT values.

Table 1
Evaluation of each paired MSA used for predicting the eSRK–SP11 complex.

	mean pDockQ[a]	mean rmsd [Å][a]	Number of best pDockQ structures
default	0.424	5.06	3
seq	0.623	3.39	24
cys-seq	0.616	3.75	24
st	0.591	3.59	15
st2	0.600	3.96	32

[a] average of the 98 predicted models

Notably, the loop between β2–β3 of class-II SP11s interacted with that of another SP11 molecule in the class-II complex models (Fig. 6**F**). Moreover, unlike that observed in the class-I SP11 models such as *Br* S_8 and *Br* S_9, the α-helix in class-II did not interact with HV III of the corresponding eSRKs (Fig. S9), indicating that the HV III of eSRK of class-II does not contribute to the ΔG_{bind}.

The high similarity of the backbone structure of the predicted SP11 and eSRK–SP11 complex models in the class-II models, except for *Rr* S_6, and the well-conserved amino acids consisting of the β3 strand of SP11 suggest that the self/nonself-discrimination within the class II is achieved by the other binding interfaces. Structural comparison of the predicted eSRK–SP11 complexes of *Br* S_{44} and *Br* S_{60} suggested that the slightly bulky Glu58 and Lys61 of S_{60}-SP11 are located in a manner so as not to interfere with Trp292 and Gly293 of S_{60}-eSRK, whereas in the S_{44} model, Ala58 and Thr61 of SP11 and bulky Arg291 and Asp292 of eSRK are located at the same position (Fig. 6**G**). According to the sequence alignment of the class-II eSRKs and SP11s (Fig. 3**H**), only S_{44} exhibited an inversion of bulkiness, suggesting that it plays a key role in the self/nonself-discrimination in class II and in the N-terminal accessorial variable region.

Interestingly, the eSRK of *Rr* S_6 lacks the four-residue FLNQ insertion characteristic of the class-II HV-I region, whereas its SP11 displayed a class-II-like sequence and predicted structure (Fig. 6**D**). The pDockQ score of the predicted *Rr* S_6 complex model was not highly plausible (pDockQ = 0.322) compared with that of the other class-II predicted models (mean pDockQ = 0.588). Hence, we independently superimposed the predicted *Rr* S_6 eSRK and SP11 models onto the predicted *Br* S_{29} complex structure to generate a new complex model equivalent to the other class-II complex models showing high pDockQ. Using the model as the initial coordinate of MD simulations, the MM–GBSA analysis demonstrated that the calculated ΔG_{bind} between the eSRK and SP11 models was negative (–36.3 kcal/mol) and identified the residues contributing to the binding. Analyzing *Rr* S_6 showed that the Tyr residue at position 56 of SP11 did not contribute to the ΔG_{bind} (Fig. 6**C**). Taken together, these observations suggest that binding and self-recognition in class II haplotypes are predominantly mediated by the interaction between the α-helix of SP11 and the HV-II/N-terminal accessorial variable region of eSRK. Additionally, the interaction between the Tyr residue at position 56 of SP11 and the FLNQ insertion of eSRK augments the affinity.

Finally, we also evaluated the five haplotypes showing the different binding mode using MD simulations. The *Ah* S_{28}, *Al* S_6, *Al* S_{14}, and *Al* S_a complex models exhibited ΔG_{bind} values of –139, –88.3, –60.3, and –98.7 kcal/mol, respectively, indicating that they were plausibly modeled and their cognate eSRK and SP11 molecules were tightly bound. In contrast, *Al* S_{18} showed ΔG_{bind} = –12.6 kcal/mol, suggesting that the model was partially inaccurate. Next, we counted the number of atomic contacts between the eSRK and SP11 molecules in the modeled *Ah* S_{28} and compared it with that of the model based on the *Br* S_9 crystal

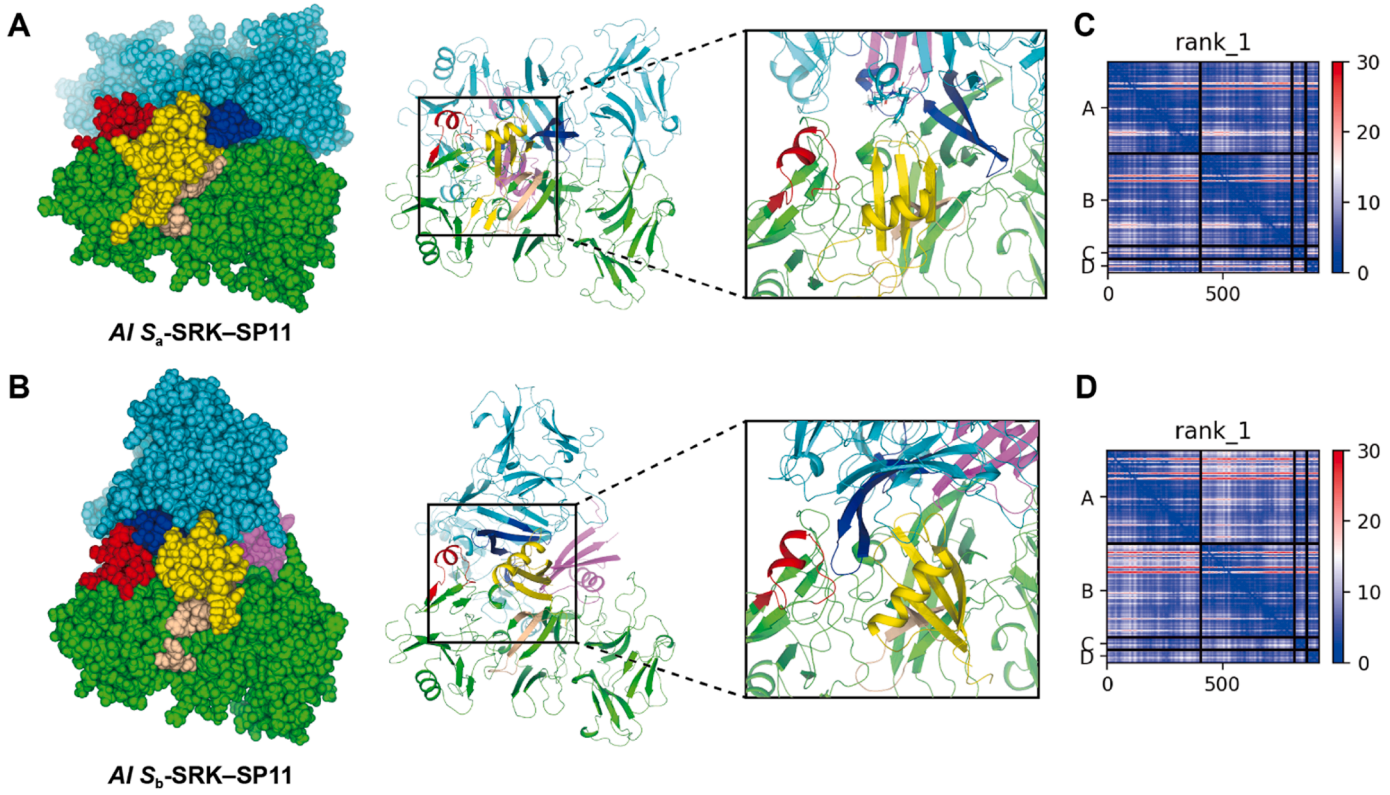


Fig. 5. Predicted binding mode for some *Al* or *Ah* SRK–SP11 complexes. (A–B) Predicted *Al* S_a (A) and *Al* S_b (B) complex models using ColabFold. Two SRK monomers are colored in green and cyan, and two SP11s are in yellow and purple. The position and orientation of *Al* S_a- and *Al* S_b- SRK chain A (green) are aligned. The HV regions I, II, and III of SRK are colored in light brown, deep blue, and red, respectively. HV-I and III belong to the same SRK chain, whereas HV-II belongs to the other chain. Sphere and cartoon representations are shown in the left and center panels, respectively. A close-up view of the binding interface between SRK and SP11 are shown in the right panel. (C–D) The predicted aligned error (PAE) for *Al* S_a (C) and *Al* S_b (D). Chains A–B and C–D represent SRK and SP11, respectively.

structure geometry [86]. The *Ah* S_{28} SRK-SP11 complex model showed a lower number of atomic contacts (1251 contacts) than that of the previously modeled one (1395 contacts) but higher than those of the *Br* S_8 or *Br* S_9 crystal structures (1180 and 1159 contacts, respectively, Table S1). Moreover, the Arg277 and Glu279 of chain A in the *Ah* S_{28}-eSRK model formed salt bridges with those in chain B (Fig. 7**A**). Additionally, Arg218 of the eSRK model formed an H-bond network with Asp288/Glu293 of the other chain of the eSRK and Arg68 of the cognate SP11 (Fig. 7**B**). Although these interactions were not observed in the previous model, our contact number analysis and ΔG_{bind} calculation suggest that the different binding mode is also probable.

4. Discussion

In this study, we demonstrated that SP11 proteins from various species could be predicted with high pLDDT using ColabFold with manually curated and custom MSAs. We have also presented eSRK–SP11 complex models using paired eSRK and SP11 sequences originating from the same *S*-haplotype. The difficulty encountered in predicting the SP11 defensin-like domain structure using the original AlphaFold2/ColabFold framework stems from alignment errors pertaining to the presence of eight or more cysteine residues within its approximately 60-residue sequence. Here, we circumvented this issue by employing four custom MSAs grounded in the assumption that Cys residues form disulfide bonds with fixed connectivity. Lastly, a comprehensive analysis of the binding mechanism between eSRK and SP11s led to the discovery of novel specificity-determining regions of SRK located at the N- and C-terminal accessorial variable regions. We expect that the computational results will provide significant clues to a comprehensive understanding of the self/nonself-discrimination mechanism in Brassicaceae, which has been deemed challenging because of the polymorphism in sequence and structure.

The accuracy of AlphaFold2 relies on the extraction of coevolutionary information from the input primary sequence and its corresponding MSA generated in the calculation pipeline. [38] The Evoformer block of AlphaFold2, inspired by direct coupling analysis (DCA) [87–91], is tasked with predicting the inter-residue distances in the input sequence by iteratively refining the spatial coordinates and evolutionary information within the MSA. Importantly, the accuracy significantly drops if the median per-residue N_{eff} is less than approximately 30 but stabilizes when N_{eff} exceeds 100. Consequently, the quantity and quality of MSAs play an important role in the accuracy of the model. However, obtaining an accurate sequence alignment is difficult for small cysteine-rich proteins with few homologs in the sequence database. Our predictions for SP11s using ColabFold and the various MSAs (Fig. 1**B**) also confirm the significance of N_{eff}. Additionally, the improved model accuracy could be attributed to the rigorous preparation of the MSA, which comprises 98 pairs of coevolving eSRK and SP11 amino acid sequences along with the meticulously corrected cysteine residue positions. Overall, the eSRK–SP11 complex prediction for *Brassica* and *Raphanus* using our custom MSAs showed high pLDDT and pDockQ values, whereas those for some *S* haplotypes of *A. halleri* and *A. lyrata* were low. This result may be attributed to the small number of sequences that are closely related to the genus *Arabidopsis* in the sequence dataset.

Our MM–GBSA analysis revealed that the calculated ΔG_{bind} values were largely negative for 57 cognate eSRK-SP11 pairs, indicating that their predicted models were physically plausible. We performed the analysis for cognate pairs in this study because the number of non-cognate pairs is substantially large. However, our models provide a reasonable explanation for interspecific pairs exhibiting the same recognition specificities such as in *Bo* S_7–*Br* S_{46}, *Bo* S_{32}–*Br* S_8, *Bo* S_{24}–*Br*

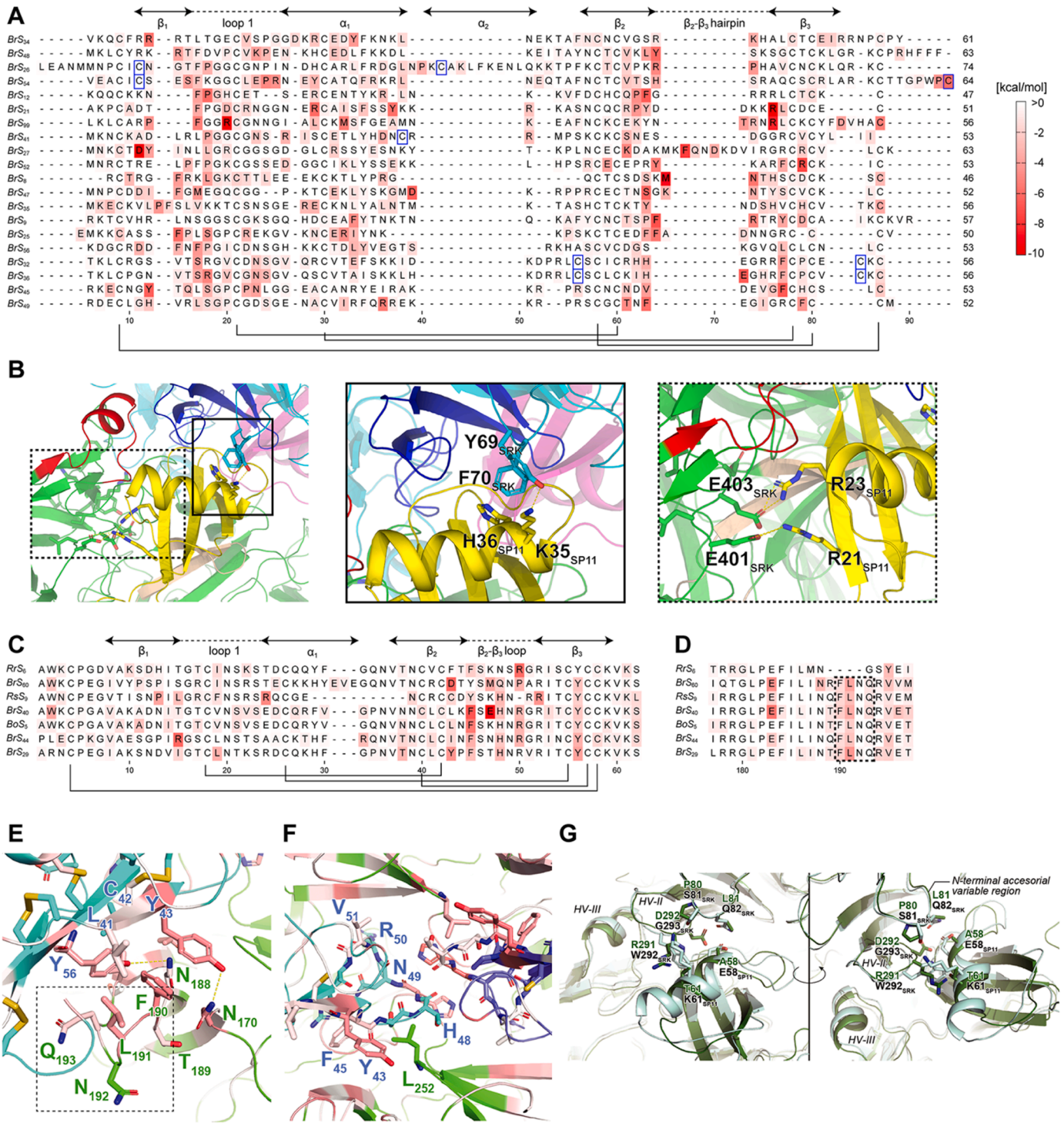


Fig. 6. Molecular mechanics–generalized born surface area (MM–GBSA) analysis using the predicted structures of SRK–SP11 complexes. (A) *Br* class-I SP11 sequences with heat maps indicating energy contributions from each residue. Cys residue pairs forming disulfide bonds are shown in black lines. Extra Cys residues observed in class-I SP11s are highlighted in blue squares. (B) N- and C-terminal accessorial variable regions of eSRK in the *Br* S_{48} SRK–SP11 complex model. Close-up views of the N- and C-terminal regions are depicted in the middle (solid line) and right (dashed line) panels, respectively. The color scheme of the protein model is the same as in Fig. 1. (C) The heat map of *Br* class-II SP11 sequences. (D) The heatmaps of HV-I of class-II ectodomain SRKs (eSRKs). The FLNQ insertion are indicated by the dashed lines. (E–F) Close-up view of the binding interface around the eSRK HV-I region (E) and the SP11 β2-β3 loop (F) in the predicted *Br* S_{29} model. eSRK and SP11 are colored as cyan and green, respectively. Residues at the interface are colored according to their per-residue ΔG_{bind} values. Subscripts next to the one-letter residue name indicate the residue number in the sequence alignment. (G) Superposition of predicted *Br* S_{44} (green) and S_{60} SRK–SP11 (pale blue) complex models.

S_{36}, *Bo* S_{12}–*Br* S_{47}, and *Bo* S_{64}–*Br* S_{41} pairs [68,92]. Their SRK and SP11 sequences have a > 92% sequence identity for each pair and superposition of their models showed that the mutated residues in each pair were not located around the eSRK–SP11 binding interfaces (Fig. S10).

Apart from the studies investigating the crystal structures of *Br* S_8 or S_9, few in vitro experimental studies in the past two decades have directly investigated the eSRK–SP11 molecular interactions because of the difficulty in expressing these proteins under laboratory conditions.

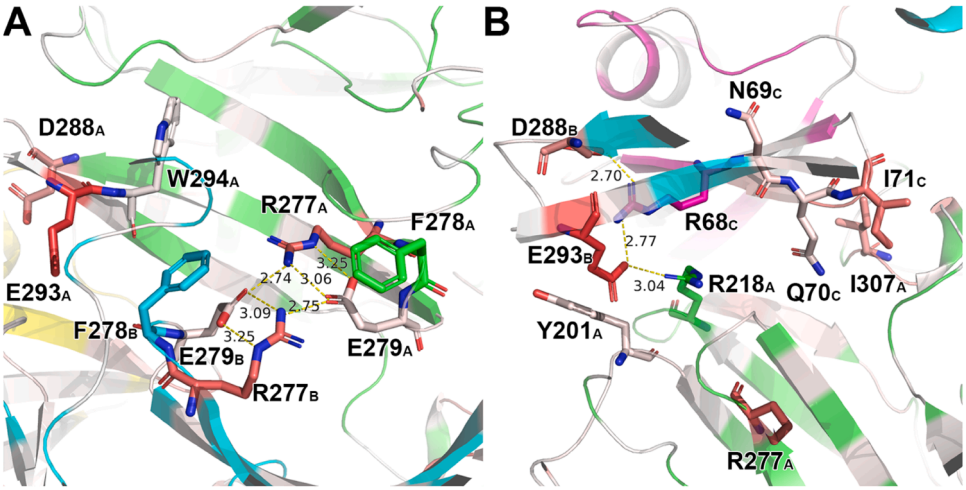


Fig. 7. The binding interface of *Ah* S_{28} eSRK–SP11 complex model. Chain A (green) and B (cyan) belong to eSRK and Chain C (purple) and D (yellow) to SP11. The dashed line represents hydrogen bonds. Bond lengths are shown in angstroms. Residues are colored according to their per-residue ΔG_{bind} values as in Fig. 6. The residue numbers are labeled in conformance with the GenBank database. (A) The interface between the two eSRK molecules. (B) eSRK–SP11 interface.

Instead, the interaction has been analyzed through an in vivo assay using transgenic plants with mutation in their HV I-III regions. Boggs et al. constructed an eSRK chimera: eSRKa(7)a. Its *Al* S_a HV I-III regions were replaced with those of *Capsella grandiflora* (*Cg*) S_7, which is 77% similar to *Al* S_a. They demonstrated that the transgenic plant expressing eSRKa (7)a showed an incompatibility response to pollen expressing *Cg* S_7-SP11. Additionally, they identified six single mutations of eSRKa(7)a that disrupt its function. We performed the prediction for *Cg* eSRK–SP11 to validate whether the results and our predicted models were consistent (Fig. S11 **and Supplementary File 4**). The predicted complex model exhibited almost the same binding mode as that of *Al* S_a, and the six residues were located at the eSRK–SP11 interface. Moreover, the models also explained the adverse impact of each single mutation on the interaction. The K213M and Y301T mutations in eSRKa(7)a were predicted to disrupt the interactions formed by Thr66 and Asp69 in *Cg* S_7-SP11, respectively, and the I217R mutant was predicted to interfere with Arg34 in *Cg* S_7-SP11. Moreover, the backbone structure of residues 294–300, which is derived from *Cg* S_7, was predicted to differ from *Al* S_a because of the exceptional phi torsion angle of Pro294. Thus, the P294S, X(gap)298 A, and D300E mutations are predicted to alter the main chain structure and thereby change the incompatibility response. Meanwhile, a weakened response caused by the L218V mutation may be attributed to a slight decrease in binding affinity between eSRKa(7)a and *Cg* S_7-SP11 owing to its smaller volume, although the model could not represent the effect well. Unfortunately, we could not examine the other chimera, eSRK16(25)16, in which they replaced the HV I-III regions of *Al* S_{16} with those from *Al* S_{25}, due to failure in predicting their complex models.

Another crucial factor that needs to be mentioned is the relevance of eSRK–SP11 binding modes and the phylogenetic characterization for the allelic diversity of the *S*-locus in *Arabidopsis* and *Brassica* species. Prigoda et al. [93] and Goubet et al. [58] have suggested that *Ah* S_{28}, *Al* S_6, *Al* S_{14}, and *Al* S_{18} belong to class B/II, and *Al* S_a (S_{13}) is distinctly classified as class A3/III. Additionally, *Ah* S_{12}, *Ah* S_{20}, *Ah* S_{32}, *Al* S_b (S_{20}), and potentially *Al* S_{36} (**Supplementary File 5**) have been embedded within class A2/IV. Notably, the different binding mode observed in the complex models of *Ah* S_{28}, *Al* S_6, *Al* S_{14}, *Al* S_{18}, and *Al* S_a correlate with the phylogenetic or dominance class, class B/II or A3/III, for SRK in *Arabidopsis* [93]. These complex models exhibited large negative free energy changes, except for *Al* S_{18}, and are thus considered plausible (Fig. 5**A and** Fig. S12). Furthermore, considering that both class-I and class-II *Brassica* SRK alleles are embedded within the class A2/IV in *Arabidopsis*, a correlation may exist between the eSRK–SP11 binding modes and the phylogenetic/dominance classes in Brassicaceae. However, owing to the limited availability of paired *Arabidopsis* SRK/SP11 sequences in the sequence database, we were unable to investigate this correlation in depth. Nevertheless, performing protein crystallization for the proteins associated with these allelic classes may be crucial to gain further insights into the evolution–structure relationship.

We have thus demonstrated that experimentally curated homologous sequences and additional manipulations to align the Cys residues yielded more plausible protein models than those predicted using the original AlphaFold2. Meanwhile, the improvement in structure prediction for eSRK–SP11 complexes using paired MSA was largely attributed to the architecture of AlphaFold2/ColabFold to consider the coevolutionary signals embedded in the MSA. This implies that as more annotated sequence pairs of eSRK and SP11 become available in the future, their prediction accuracy will enhance. We anticipate that a collaborative effort between experimental and computational biology will unravel the complexities of self-incompatibility in Brassicaceae.

CRediT authorship contribution statement

Tomoki Sawa: Methodology, Investigation, Data curation, Writing – original draft. **Yoshitaka Moriwaki:** Conceptualization, Methodology, Investigation, Validation, Data Curation, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Hanting Jiang:** Investigation. **Kohji Murase:** Conceptualization, Resources, Writing – review & editing. **Seiji Takayama:** Writing – review & editing, Supervision. **Kentaro Shimizu:** Supervision. **Tohru Terada:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was partially supported by "JSPS KAKENHI Grant Numbers JP21K06110" and "Research Support Project for Life Science and Drug Discovery (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Numbers JP23ama121026 (support number 4487) and JP23ama121027." This study was conducted using the TSUBAME3.0 supercomputer at Tokyo

Institute of Technology, the General Projects on supercomputer "Flow" at Information Technology Center, Nagoya University, and the FUJITSU Supercomputer PRIMEHPC FX1000 and FUJITSU Server PRIMERGY GX2570 (Wisteria/BDEC-01) at the Information Technology Center, The University of Tokyo. We would like to thank Editage (www.editage.jp) for English language editing.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.10.026.

References

[1] de Nettancourt D. Incompatibility and Incongruity in Wild and Cultivated Plants. second ed. Berlin: Springer; 2001.
[2] Takayama S, Isogai A. Self-incompatibility in plants. Annu Rev Plant Biol 2005;56:467–89.
[3] Bateman AJ. Self-incompatibility systems in angiosperms: III. Cruciferae. Heredity 1955;9:53–68.
[4] Nou S, Watanabe M, Isogai A, Hinata K. Comparison of *S*-alleles and *S*-glyeoproteins between two wild populations of *Brassica campestris* in Turkey and Japan. Sex Plant Reprod 1993;6:79–86.
[5] Ruffio-Chable V, Gaude T. *S*-haplotype polymorphism in *Brassica oleracea*. Acta Hortic 2001:257–61.
[6] Kao TH, McCubbin AG. How flowering plants discriminate between self and non-self pollen to prevent inbreeding. Proc Natl Acad Sci USA 1996;93:12059–65.
[7] Kao TH, Tsukamoto T. The molecular and genetic bases of S-RNase-based self-incompatibility. Plant Cell 2004;16:S72–83.
[8] McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, et al. Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. Nature 1989;342:955–7.
[9] Franklin-Tong N, Franklin FCH. Gametophytic self-incompatibility inhibits pollen tube growth using different mechanisms. Trends Plant Sci 2003;8:598–605.
[10] Foote HCC, Ride JP, Franklintong VE, Walker EA, Lawrence MJ, Franklin FCH. Cloning and expression of a distinctive class of self-incompatibility (*S*) gene from *Papaver rhoeas* L. Proc Natl Acad Sci USA 1994;91:2265–9.
[11] Lane MD, Lawrence MJ. The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. VII. Number *S*-alleles Species, *Hered* 1993;71:596–602.
[12] Hearn MJ, Franklin FCH, Ride JP. Identification of a membrane glycoprotein in pollen of Papaver rhoeas which binds stigmatic self-incompatibility (S-) proteins. Plant J 1996;9:467–75.
[13] Jordan ND, Kakeda K, Conner A, Ride JP, Franklin-Tong VE, Franklin FCH. S-protein mutants indicate a functional role for SBP in the self-incompatibility reaction of *Papaver rhoeas*. Plant J 1999;20:119–25.
[14] Nasrallah ME, Wallace DH. Immunogenetics of self-incompatibility in *Brassica oleracea L.* Heredity 1967;22:519–27.
[15] Nishio T, Hinata K. Analysis of S-specific proteins in stigma of *Brassica oleracea* L. by isoelectric focusing. Heredity 1977;38:391–6.
[16] Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K. The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. Nature 2000;403:913–6.
[17] Stein JC, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB. Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. Proc Natl Acad Sci USA 1991;88:8816–20.
[18] Takayama S, Shiba H, Iwano M, Shimosato H, Che FS, Kai N, et al. The pollen determinant of self-incompatibility in *Brassica campestris*. Proc Natl Acad Sci USA 2000;97:1920–5.
[19] Schopfer CR, Nasrallah ME, Nasrallah JB. The male determinant of self-incompatibility in *Brassica*. Science 1999;286:1697–700.
[20] Shiba H, Takayama S, Iwano M, Shimosato H, Funato M, Nakagawa T, et al. A pollen coat protein, SP11/SCR, determines the pollen *S*-specificity in the self-incompatibility of *Brassica* species. Plant Physiol 2001;125:2095–103.
[21] Suzuki G, Kai N, Hirose T, Fukui K, Nishio T, Takayama S, et al. Genomic organization of the *S* locus: Identification and characterization of genes in SLG/SRK region of S^9 haplotype of *Brassica campestris* (syn. *rapa*). Genetics 1999;153:391–400.
[22] Takayama S, Shimosato H, Shiba H, Funato M, Che FS, Watanabe M, et al. Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility. Nature 2001;413:534–8.
[23] Kachroo A, Schopfer CR, Nasrallah ME, Nasrallah JB. Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. Science 2001;293:1824–6.
[24] Shimosato H, Yokota N, Shiba H, Iwano M, Entani T, Che FS, et al. Characterization of the SP11/SCR high-affinity binding site involved in self/nonself recognition in *Brassica* self-incompatibility. Plant Cell 2007;19:107–17.
[25] Cabrillac D, Cock JM, Dumas C, Gaude T. The S-locus receptor kinase is inhibited by thioredoxins and activated by pollen coat proteins. Nature 2001;410:220–3.
[26] Nasrallah JB. Stop and go signals at the stigma–pollen interface of the Brassicaceae. kiad301 Plant Physiol 2023. kiad301.
[27] Boyes DC, Nasrallah JB. Physical linkage of the *SLG* and *SRK* genes at the self-incompatibility locus of *Brassica oleracea*. Mol Gen Genet 1993;236:369–73.
[28] Naithani S, Chookajorn T, Ripoll DR, Nasrallah JB. Structural modules for receptor dimerization in the *S*-locus receptor kinase extracellular domain. Proc Natl Acad Sci USA 2007;104:12211–6.
[29] Kusaba M, Nishio T, Satta Y, Hinata K, Ockendon D. Striking sequence similarity in inter- and intra-specific comparisons of class I *SLG* alleles from *Brassica oleracea* and *Brassica campestris*: Implications for the evolution and recognition mechanism. Proc Natl Acad Sci USA 1997;94:7673–8.
[30] Sato K, Nishio T, Kimura R, Kusaba M, Suzuki T, Hatakeyama K, et al. Coevolution of the *S*-locus genes *SRK*, *SLG* and *SP11/SCR* in *Brassica oleracea* and *B. rapa*. Genetics 2002;162:931–40.
[31] Watanabe M, Ito A, Takada Y, Ninomiya C, Kakizaki T, Takahata Y, et al. Highly divergent sequences of the pollen self-incompatibility (*S*) gene in class-I *S* haplotypes of *Brassica campestris* (syn. *rapa*) L. FEBS Lett 2000;473:139–44.
[32] Shiba H, Iwano M, Entani T, Ishimoto K, Shimosato H, Che FS, et al. The dominance of alleles controlling self-incompatibility in *Brassica* pollen is regulated at the RNA level. Plant Cell 2002;14:491–504.
[33] Tarutani Y, Shiba H, Iwano M, Kakizaki T, Suzuki G, Watanabe M, et al. Trans-acting small RNA determines dominance relationships in *Brassica* self-incompatibility. Nature 2010;466:983–6.
[34] Yasuda S, Wada Y, Kakizaki T, Tarutani Y, Miura-Uno E, Murase K, et al. A complex dominance hierarchy is controlled by polymorphism of small RNAs and their targets. Nat Plants 2017;3:16206.
[35] Ma R, Han ZF, Hu ZH, Lin GZ, Gong XQ, Zhang HQ, et al. Structural basis for specific self-incompatibility response in *Brassica*. Cell Res 2016;26:1320–9.
[36] Murase K, Moriwaki Y, Mori T, Liu X, Masaka C, Takada Y, et al. Mechanism of self/nonself-discrimination in *Brassica* self-incompatibility. Nat Commun 2020;11:4916.
[37] Ockendon DJ. The *S*-allele collection of *Brassica oleracea*. Acta Hortic 2000;539:25–30.
[38] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9.
[39] Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. 2004.463034 bioRxiv 2022;2021(2010). 2004.463034.
[40] Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinforma 2010;11:431.
[41] Remmert M, Biegert A, Hauser A, Soding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods 2012;9:173–5.
[42] Steinegger M, Meier M, Mirdita M, Vohringer H, Haunsberger SJ, Soding J. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinforma 2019;20:473.
[43] Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods 2022;19:679–82.
[44] Steinegger M, Soding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 2017;35:1026–8.
[45] Mirdita M, Steinegger M, Soding J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. Bioinformatics 2019;35:2856–8.
[46] Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 2022;50:D439–44.
[47] Glavin TL, Goring DR, Schafer U, Rothstein SJ. Features of the extracellular domain of the *S*-locus receptor kinase from *Brassica*. Mol Gen Genet 1994;244:630–7.
[48] Fukai E, Fujimoto R, Nishio T. Genomic organization of the *S* core region and the *S* flanking regions of a class-II *S* haplotype in *Brassica rapa*. Mol Genet Genom 2003;269:361–9.
[49] Vanoosthuyse V, Miege C, Dumas C, Cock JM. Two large *Arabidopsis thaliana* gene families are homologous to the *Brassica* gene superfamily that encodes pollen coat proteins and the male component of the self-incompatibility response. Plant Mol Biol 2001;46:17–34.
[50] Boggs NA, Dwyer KG, Shah P, McCuuoch AA, Bechsgaard J, Schierup MH, et al. Expression of distinct self-incompatibility specificities in *Arabidopsis thaliana*. Genetics 2009;182:1313–21.
[51] Delorme V, Giranton JL, Hatzfeld Y, Friry A, Heizmann P, Ariza MJ, et al. Characterization of the *S* locus genes, *SLG* and *SRK*, of the *Brassica* S_3 haplotype: identification of a membrane-localized protein encoded by the *S* locus receptor kinase gene. Plant J 1995;7:429–40.
[52] Suzuki T, Kusaba M, Matsushita M, Okazaki K, Nishio T. Characterization of *Brassica S*-haplotypes lacking *S*-locus glycoprotein. FEBS Lett 2000;482:102–8.
[53] Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, Nasrallah ME. Self-incompatibility in the genus *Arabidopsis*: Characterization of the *S* locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. Plant Cell 2001;13:627–43.
[54] Takuno S, Oikawa E, Kitashiba H, Nishio T. Assessment of genetic diversity of accessions in Brassicaceae genetic resources by frequency distribution analysis of *S* haplotypes. Theor Appl Genet 2010;120:1129–38.
[55] Kusaba M, Matsushita M, Okazaki K, Satta Y, Nishio T. Sequence and structural diversity of the *S* locus genes from different lines with the same self-recognition specificities in *Brassica oleracea*. Genetics 2000;154:413–20.
[56] Yamakawa S, Watanabe M, Hinata K, Suzuki A, Isogai A. The Sequences of *S*-Receptor Kinases (SRK) Involved in Self-incompatibility and Their Homologies to *S*-Locus Glycoproteins of *Brassica campestris*. Biosci, Biotechnol, Biochem 1995;59:161–2.
[57] Kumar V, Trick M. Expression of the *S*-locus receptor kinase multigene family in *Brassica oleracea*. Plant J 1994;6:807–13.

[58] Goubet PM, Berges H, Bellec A, Prat E, Helmstetter N, Mangenot S, et al. Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in *Arabidopsis*. PLoS Genet 2012;8:e1002495.
[59] Kakizaki T, Takada Y, Fujioka T, Suzuki G, Satta Y, Shiba H, et al. Comparative analysis of the *S*-intergenic region in class-II *S* haplotypes of self-incompatible *Brassica rapa* (syn. *campestris*). Genes Genet Syst 2006;81:63–7.
[60] Cabrillac D, Delorme V, Garin J, Ruffio-Chable V, Giranton JL, Dumas C, et al. The S_{15} self-Incompatibility haplotype in *Brassica oleracea* includes three *S* gene family members expressed in stigmas. Plant Cell 1999;11:971–86.
[61] Bechsgaard JS, Castric V, Vekemans X, Schierup MH. The transition to self-compatibility in *Arabidopsis thaliana* and evolution within *S*-haplotypes over 10 Myr. Mol Biol Evol 2006;23:1741–50.
[62] Fujimoto R, Okazaki K, Fukai E, Kusaba M, Nishio T. Comparison of the genome structure of the self-incompatibility (*S*) locus in interspecific pairs of *S* haplotypes. Genetics 2006;173:1157–67.
[63] Dwyer KG, Berger MT, Ahmed R, Hritzo MK, McCulloch AA, Price MJ, et al. Molecular characterization and evolution of self-Incompatibility genes in *Arabidopsis thaliana*: The case of the *Sc* haplotype. Genetics 2013;193:985–94.
[64] Kusaba M, Nishio T. Comparative analysis of *S* haplotypes with very similar *SLG* alleles in *Brassica rapa* and *Brassica oleracea*. Plant J 1999;17:83–91.
[65] Hatakeyama K, Takasaki T, Watanabe M, Hinata K. Molecular characterization of *S* locus genes, *SLG* and *SRK*, in a pollen-recessive self-incompatibility haplotype of *Brassica rapa* L. Genetics 1998;149:1587–97.
[66] Okamoto S, Odashima M, Fujimoto R, Sato Y, Kitashiba H, Nishio T. Self-compatibility in *Brassica napus* is caused by independent mutations in *S*-locus genes. Plant J 2007;50:391–400.
[67] Suzuki G, Watanabe M, Isogai A, Hinata K. Highly conserved 5'-flanking regions of two self-incompatibility genes, SLG^9 and SRK^9. Gene 1997;191:123–6.
[68] Sato Y, Fujimoto R, Toriyama K, Nishio T. Commonality of self-recognition specificity of *S* haplotypes between *Brassica oleracea* and *Brassica rapa*. Plant Mol Biol 2003;52:617–26.
[69] Takuno S, Fujimoto R, Sugimura T, Sato K, Okamoto S, Zhang SL, et al. Effects of recombination on hitchhiking diversity in the *Brassica* self-incompatibility locus complex. Genetics 2007;177:949–58.
[70] Boggs NA, Nasrallah JB, Nasrallah ME. Independent *S*-Locus mutations caused self-fertility in *Arabidopsis thaliana*. PLoS Genet 2009;5:e1000426.
[71] Mirdita M, von den Driesch L, Galiez C, Martin MJ, Soding J, Steinegger M. Uniclust databases of clustered and deeply annotated protein sequences and alignments. Nucleic Acids Res 2017;45:D170–6.
[72] Shafee TMA, Robinson AJ, van der Weerden N, Anderson MA. Structural homology guided alignment of cysteine rich proteins. Springerplus 2016;5:27.
[73] Holm L. Dali server: structural unification of protein families. Nucleic Acids Res 2022;50:W210–5.
[74] Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Res 2010;38:W545–9.
[75] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011;7:539.
[76] Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci 2018;27:135–45.
[77] Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. MUSTANG: A multiple structural alignment algorithm. Protein: Struct Funct Bioinform 2006;64:559–74.
[78] Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics 2013;29:2722–8.
[79] Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. Nat Commun 2022;13:1265.
[80] Kollman PA, Massova I, Reyes C, Kuhn B, Huo SH, Chong L, et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–97.
[81] D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D. S. Cerutti, T.E.I. Cheatham, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, K. P.A, AMBER 22 2022 University of California, San Francisco.
[82] Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migues AN, et al. ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energies in solution. J Chem Theory Comput 2020;16:528–52.
[83] Izadi S, Anandakrishnan R, Onufriev AV. Building water models: a different approach. J Phys Chem Lett 2014;5:3863–71.
[84] Roe DR, Okur A, Wickstrom L, Hornak V, Simmerling C. Secondary structure bias in generalized born solvent models: comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. J Phys Chem B 2007;111:1846–57.
[85] Shafee TMA, Lay FT, Hulett MD, Anderson MA. The defensins consist of two independent, convergent protein superfamilies. Mol Biol Evol 2016;33:2345–56.
[86] Chantreau M, Poux C, Lensink MF, Brysbaert G, Vekemans X, Castric V. Asymmetrical diversification of the receptor-ligand interaction controlling self-incompatibility in *Arabidopsis*. eLife 2019;8:e50253.
[87] Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing. Proc Natl Acad Sci USA 2009;106:67–72.
[88] Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci USA 2011;108:E1293–301.
[89] Balakrishnan S, Kamisetty H, Carbonell JG, Lee SI, Langmead CJ. Learning generative models for protein fold families. Protein: Struct Funct Bioinform 2011; 79:1061–78.
[90] Ekeberg M, Lovkvist C, Lan YH, Weigt M, Aurell E. Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. Phys Rev E: Stat Phys, Plasmas, Fluids 2013;87:012707.
[91] Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc Natl Acad Sci USA 2013;110:15674–9.
[92] Kimura R, Sato K, Fujimoto R, Nishio T. Recognition specificity of self-incompatibility maintained after the divergence of *Brassica oleracea* and *Brassica rapa*. Plant J 2002;29:215–23.
[93] Prigoda NL, Nassuth A, Mable BK. Phenotypic and genotypic expression of self-incompatibility haplotypes in *Arabidopsis lyrata* suggests unique origin of alleles in different dominance classes. Mol Biol Evol 2005;22:1609–20.
[94] Wu TQ, Hou J, Adhikari B, Cheng JL. Analysis of several key factors influencing deep learning-based inter-residue contact prediction. Bioinformatics 2020;36: 1091–8.