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

Tomoki Sawa[†], Yoshitaka Moriwaki[†,§,*], Hanting Jiang[†], Kohji Murase[‡], Seiji Takayama[‡], Kentaro Shimizu[†,§], and Tohru Terada[†,§]

[†]Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan.

[§]Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan.

[‡]Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan.

[*]Corresponding author. E-mail: moriwaki@bi.a.u-tokyo.ac.jp

**ABSTRACT:** Plants employ self-incompatibility (SI) to promote cross-fertilization. In Brassicaceae, this process is regulated by the complex formation of 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 demonstrated that only the SRK ectodomain (eSRK) and SP11 pairs derived from the same *S* haplotype exhibit high binding free energy based on the crystal structures of two eSRK–SP11 complexes in *Brassica rapa* $S_8$ and $S_9$ haplotypes and nine computationally predicted complex models. However, predicting the eSRK–SP11 complex structures for the other 100+ *S* haplotypes and genera remains difficult because of the SP11 polymorphism in sequence and structure. Although protein structure prediction using AlphaFold2 exhibits considerably high accuracy for most protein monomers and complexes, 46% of predicted SP11 structures we tested showed per-residue mean pLDDT < 75. Here, we demonstrate that the use of curated multiple sequence alignment (MSA) for cysteine-rich proteins significantly improves the 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, these computational results provide insights into the haplotype-specific eSRK–SP11 binding modes in Brassicaceae and could be adopted as a general method for obtaining more accurate models for other cysteine-rich proteins.

## 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, 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 in *B. oleracea* (*Bo*) and *B. rapa* (*Br*)[16, 17]. Furthermore, *S*-locus protein 11 (SP11, also known as *S*-locus cysteine-rich protein (SCR)) has been identified as a determinant present on the pollen side[18-21]. SRK, located on the plasma membrane of papilla cells, specifically binds to its cognate SP11 released from the pollen surface, triggering the self-phosphorylation of the kinase domain inside the cell and resulting in the rejection of fertilization[22-25]. The alleles of *SLG*, *SRK*, and *SP11* are tightly linked at the *S*-locus and transmitted to the progeny as a single set called *S* haplotype[18, 26].

The amino acid sequences of SRK and SP11 were significantly different among the *S* haplotypes. The ectodomain of SRK (eSRK) shares two lectin domains with an EGF-like domain and a PAN domain among the haplotypes[27] in addition to three hypervariable (HV I–III) regions with markedly different sequence compositions[28, 29]. In contrast, SP11 is a defensin-like protein that consists of 60–70 amino acids with four or five disulfide bonds characterized by numerous insertions, deletions, and variations across haplotypes[30]. Based on sequence similarity, the *S* haplotypes

are classified into classes I and II in *Brassica* and *Raphanus*. In *Br*, class II comprises the $S_{29}$, $S_{40}$, $S_{44}$, and $S_{60}$ haplotypes[31]. In terms of pollen phenotypes, the dominant nature of class I 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*[32, 33].

Recently, two eSRK–SP11 complex crystal structures derived from *Br* $S_8$ and $S_9$ have been reported. The *Br* $S_9$-eSRK and $S_9$-SP11 complexes have revealed that the two SP11 molecules were tightly bound to the interface of the two eSRKs[34]. We have reported the crystal structures of engineered *Br* $S_8$-eSRK–$S_8$-SP11 and have presented nine complex structural models of $S_8$-, $S_9$-relative and class II haplotypes using homology modeling and accelerated molecular dynamics (MD) simulations[35]. We have also validated that a point mutation in residue of $S_{36}$-SP11 located at the interface, inferred from the $S_{36}$-eSRK–$S_{36}$-SP11 complex model, disrupted the SI reaction in the pollination bioassay[35]. 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, a comprehensive understanding of the mechanism across all *S* haplotypes remains elusive because the number of haplotypes is estimated to be more than 100 and 50 for *Br*[4] and *Bo*[36], respectively, and modeling polymorphic SP11 using accelerated MD simulations is computationally expensive.

AlphaFold2[37] and AlphaFold-Multimer (AF-Multimer)[38] released in 2021, show high accuracy in predicting both monomer and complex structures. Briefly, AlphaFold2/AF-Multimer predicts the structure using the following steps: First, multiple sequence alignment (MSA) for the input sequence is retrieved from the sequence database using jackHMMER[39] and HHblits[40], and template structures are searched in the Protein Data Bank (PDB) database using HHSearch[41]; Second, the Evoformer processes the MSA and template structures to obtain their converged evolutionary and spatial relationships (denoted as MSA and pair representations in this paper); Third, the Structure Module constructs 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 demonstrates that the number and quality of MSA largely affect their accuracy.

The great success of AlphaFold2 in the field of protein structure prediction is expected to help greatly in understanding the structure–function relationships in proteins. ColabFold[42], a derivative of AlphaFold2, offers the structure prediction through a web browser, leveraging the MMSeqs2 webserver[43, 44] to obtain an MSA file of the input sequence. Moreover, the predicted monomeric structures of almost all proteins recorded in the UniProt database became available in the AlphaFold Protein Structure Database (AlphaFold DB)[45] after July 2022. However, although the per-residue model confidence (a pLDDT metric)[37] is sufficiently high for the eSRK model structures registered in AlphaFold DB, that of SP11 is relatively low, which correlates with the limited number of homologous sequences in SP11. Furthermore, the AlphaFold DB does not provide structural information on any protein complexes 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 annotated in previous experimental studies to overcome the drawbacks of structure prediction with AlphaFold2. Moreover, we comprehensively explore the molecular recognition mechanism between eSRK and SP11 at the residue level based on predicted complex models to identify the presence of variable regions other than HV I–III that may contribute to self/nonself-discrimination in all haplotypes.

## RESULTS

### Predicted eSRK and SP11 structures 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 (**Figure 1A**). Of the 98 haplotypes, 97 eSRK predicted structures exhibited an average pLDDT of 75 or higher, whereas only 53 predicted SP11 structures did. 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 short amino acid length and many diverse insertions and deletions in the defensin-like domain of SP11 in MSA made the prediction difficult. The relatively low pLDDT observed in *Rs* $S_{22}$-, $S_{26}$-, and $S_{30}$-SRK were due to the partial deficiency of the N- or C-terminal sequences (**Figure S1**). **Figure 1B** shows the correlation between $N_{\mathrm{eff}}$ and the average pLDDT of the structure predicted using the default MSA. The average pLDDT increased as $N_{\mathrm{eff}}$ approached 100, which is consistent with the discussion of AlphaFold2[37]. However, we observed five outlier proteins, namely, *Br* $S_{12}$, *Br* $S_{33}$, *Rs* $S_{23}$, *Rs* $S_{31}$, and *Bo* $S_{11}$ with low prediction accuracy for haplotypes and $N_{\mathrm{eff}}$ greater than 500 because their MSAs were inadequately obtained for full-length SP11 and their sequence alignments were imprecisely executed with many fragmentary homologous sequences.

### Custom MSAs improved SP11 models

ColabFold allows us to specify another MSA file as input instead of the default MSA. To obtain improved SP11 model structures, we prepared four types of MSA, denoted as "cys-seq", "seq", "st", and "st2", each consisting of 98 SP11 sequences that were experimentally annotated (see Methods). All four custom MSAs yielded SP11 models with a pLDDT equivalent to or higher than that of the default MSA (**Figure 1C and D**). For 66 of the 98 SP11s, the prediction accuracy was improved using one of the four MSAs. The median pLDDT increased to 84.0 by choosing the best model of the four MSAs, and the number of SP11s showing pLDDT > 75 increased from 53 to 81. In particular, the *Br* $S_{12}$-SP11 model predicted using cys-seq MSA showed the largest improvement (46.0 to 89.5 in the mean pLDDT; **Figure 1E**). Similarly, the prediction with custom MSAs also improved for *Br* $S_{33}$, *Bo* $S_{11}$, *Rs* $S_{23}$, and *Rs* $S_{31}$ (**Figure S2**).

Although the four MSAs improved pLDDTs 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 (**Figure 3J and Figure S3**). Similarly, the pLDDTs of *Br* $S_{46}$, $S_{61}$, *Bo* $S_7$, $S_{13}$, and $S_{13b}$ were below 70 despite their sequences being close to *Br* $S_8$, the crystal structure of which was determined. All pLDDTs for the predicted

models obtained using the default and four MSAs are provided in **Supplementary File 2**.

## Structural novelty in the predicted SP11 proteins

Our predicted models with high pLDDT values enable us to discuss 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[46], 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 (**Figure 2A–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 correctly predicted in our previous study using homology modeling and accelerated MD simulations[35] (**Figure S4**). The predicted SP11 models of *Bo* $S_{24}$ and *Bo* $S_{68}$ also exhibited the same connectivity and closely resembled *Br* $S_{36}$ and *Br* $S_{32}$, respectively, because their sequence similarities were very high (**Figure S5**). 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 (**Figure 2C–E**). Notably, SP11 of *Rs* $S_{31}$ formed a β4 strand at its C-terminus (**Figure 2E**). Third, *Br* $S_{26}$-SP11 showed the 1-10, 2-5, 3-7, 4-8, and 6-9 connectivity (**Figure 2F**). It possesses two helices between the β1 and β2 strands, although the second helix is not involved in the interaction with its corresponding eSRK (**Figure S6**).

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

Of the class-II 10 *S* haplotypes, *Br* $S_{29}$, $S_{40}$, $S_{44}$, $S_{60}$, *Bo* $S_5$, *Rr* $S_6$, and *Rs* $S_9$ exhibited a mean per-residue pLDDT of 75 or higher (**Figure 3A–G, middle**), and their backbone structures resembled each other (**Figure 3H**). The prediction for the remaining three haplotypes, *Rs* $S_{11}$, $S_{15}$, and $S_{26}$, 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 correctly predicted structures improved their pLDDT values to > 87 (**Figure 3A–G, right**). The predicted class-II structures had a characteristic β-bridge secondary structure after the first β-strand (**Figure 3I and J**). The α-helix of *Br* $S_{60}$ was longer than that of these haplotypes by two residues (**Figure 3D**). Remarkably, only the *Rs* $S_9$ model lacked the α-helix between the β1 and β2 strands among the 98 SP11s (**Figure 3G**) regardless of its high mean pLDDT (91.7). All seven predicted structures showed substantially higher pLDDT values than of those presented in the AlphaFold DB version 2022-11-01 (**Figure 3A–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 the class-II, and to the inappropriate alignment of cysteine-rich protein fragments during the construction of the MSA.

## Improved eSRK–SP11 complex prediction using ColabFold and paired MSAs

We also investigated the effect of the four MSAs on the modeling of eSRK–SP11 complexes. We used three criteria, the pTM-score[37], pDockQ score[47] and a root-mean-square-deviation (rmsd) for the crystal structure of the *Br* $S_8$ eSRK–SP11 complex (PDB:6KYW) to assess the plausibility of the predicted eSRK–SP11 complex models. Of the 98 predicted complex models with their default MSAs, only 41 showed pDockQ > 0.6. However, any of the four MSAs paired with the MSA of the eSRK enabled 80 models to meet the criteria. (**Figure 4A**). For example, one of the custom MSAs for *Br* $S_{34}$ and $S_{55}$ corrected the disulfide bond connectivity of their SP11s, allowing SP11 to bind to the interfaces of their eSRK dimers (**Figure 4C and D**). In addition, the most effective MSAs for complex prediction differed depending on the species and *S* haplotype (**Figure 4C–F, Table 1**).

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 $S_9$. The complex predictions for *Rs* $S_{11}$, $S_{15}$, $S_{26}$ were not successful owing to their poor SP11 models as described earlier. Remarkably, among the prediction wfor *Ah* and *Al*, *Ah* $S_{28}$, *Al* $S_6$, $S_{14}$, $S_{18}$, and $S_a$ showed rmsd > 10 Å despite pDockQ > 0.50, although their monomer structures were predicted with high pLDDT values (**Figure 4B**). 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 (**Figure 5A**), whereas *Ah* $S_{12}$, $S_{20}$, $S_{32}$, *Al* $S_{36}$, and $S_b$ showed a complex mode similar to that of the *Br* $S_8$ or $S_9$ crystal structures, in which HV-II was in direct contact with HV-III (**Figure 5B**). The predicted aligned error (PAE) indicates the unreliability of the residue–residue relationship of the predicted AlphaFold2 model[37]. PAE was generally low for both models (**Figure 5C and D**), suggesting that these predicted models were both plausible. Interestingly, these structural groups were consistent with the phylogenetic tree of the eSRK sequences (**Supplementary File 3**), suggesting that some *Al* and *Ah* eSRKs sequences have evolved to generate different self/nonself-discrimination mechanisms.

**Table 1. Evaluation of each MSA used to predict the eSRK–SP11 complex.**

| | mean pDockQ | mean rmsd [Å] | Num. of best pDockQ structure |
|---|---|---|---|
| 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 |

pDockQ and rmsd represent the average of the 98 predicted models.

## Comprehensive analysis of the self/nonself-discrimination mechanism between eSRKs and SP11s

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 prediction of AlphaFold2/ColabFold is statistical and not physical. Hence, we employed MM–GBSA analysis[48] to calculate the binding free energy ($\Delta G_{bind}$) for the predicted models.

Of the 98 complex models that were examined, 57 structures demonstrated $\Delta G_{bind} < -75$ kcal/mol. **Figure 6A** illustrates the per-

residue contribution to $\Delta G_{bind}$ on a sequence alignment of class-I *Br* SP11s arranged by CysBar[49] and Clustal Omega[50, 51]. In general, residues consisting of the first loop region, the α helix, and the β2-β3 hairpin exhibited large negative $\Delta G_{bind}$, which conforms with the results of our previous study[35] and the binding interfaces observed in the predicted complex models. The MM–GBSA analysis showed largely negative $\Delta G_{bind}$ values for haplotypes whose SP11 showed a characteristic secondary structure; in *Br* $S_{26}$, the contribution of $\Delta G_{bind}$ to eSRK interactions was observed near the kink between the first and second helices (**Figure S6**); 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 $\Delta G_{bind}$ in this study are shown in **Supplementary File 4**.

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 located in the N-terminal accessorial variable region interacted with the α-helix of SP11, and Glu401 and Glu403 located in the C-terminal region formed salt bridges with Arg21 and Arg23 of SP11, respectively (**Figure 6B**). 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 4**). In particular, the C-terminal variable region was previously discovered by Kusaba *et al.* but was not believed to be involved in the specificity between SRK and SP11[28]. However, the predicted models and analyses suggest that this contributes to the complex stabilization of some *S* haplotypes.

The predicted class-II complex structures presented a different binding mode from that observed for class I but shared it with the class. 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* $S_6$ (**Figure 6C–E and Figure 3H**). Notably, the loop between β2–β3 of class II SP11s interacted with that of another SP11 molecule in the class II complex models (**Figure 6F**). Moreover, unlike that observed in class-I models such as *Br* $S_8$ and $S_9$, the α-helix of class-II SP11s did not interact with HV III of the corresponding eSRKs (**Figure S7**), indicating that the HV III of eSRK of class-II does not contribute to the $\Delta G_{bind}$.

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 (**Figure 6G**). According to the sequence alignment of the class-II eSRKs and SP11s (**Figure 3H**), 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 (**Figure 6D**). 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 superimposed the predicted *Rr* $S_6$ eSRK and SP11 models onto the predicted complex structure of *Br* $S_{29}$ independently 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 $\Delta G_{bind}$ 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 $\Delta G_{bind}$ (**Figure 6C**). 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 HVII/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.

## DISCUSSION

In this study, we demonstrated that SP11 proteins from various species could be predicted with high pLDDT using manually curated and custom MSAs and ColabFold. We have also presented various eSRK–SP11 complex models using paired eSRK and SP11 sequences derived from the same *S*-locus. The difficulty in predicting the structure of the defensin-like domain of SP11 using the original AlphaFold2/ColabFold arises from the errors in the alignment of the eight or more cysteine residues present in its approximately 60-residue sequence. Here, we circumvented this issue by using the four custom MSAs based on the assumption that Cys residues form disulfide bonds with fixed connectivity. Finally, we performed a comprehensive analysis of the binding mechanism between eSRK and SP11s and discovered new specificity-determining regions called the N- and C-terminal accessorial variable regions. We believe that these observations will lead to significant advances in the understanding of the self/nonself-discrimination mechanism in Brassicaceae, which has been deemed challenging to comprehensively analyze because of polymorphisms in sequence and structure.

The high accuracy of AlphaFold2 relies on the coevolutionary information obtained from the input primary sequence and its MSA generated in the calculation pipeline.[37] Inspired by direct coupling analysis (DCA)[52-56], the Evoformer block of AlphaFold2 predicts the residue–residue distances of input sequence by continuously refining the spatial coordinates and evolutionary information within the MSA. This study showed that the accuracy of AlphaFold2 decreases substantially if the median per-residue $N_{eff}$ is less than approximately 30 but saturates at more than 100. However, for cysteine-rich proteins with few homologs in the sequence database, accurate sequence alignment becomes challenging; thus, the quantity and quality of MSAs may significantly affect model accuracy. Our prediction results for SP11s with ColabFold and the default MSA (**Figure 1B**) also support $N_{eff}$ dependence. Furthermore. the enhanced model accuracy could be attributed to the proper preparation of MSA consisting of 98 pairs of coevolving

eSRK and SP11 amino acid sequences along with their corrected cysteine residue positions.

In general, the eSRK–SP11 complex prediction for *Brassica* and *Raphanus* using our custom MSAs showed high pLDDT, pDockQ, and $\Delta G_{bind}$ values. The predicted complex models can provide a reasonable explanation for $S_7$, $S_{32}$, $S_{24}$, $S_{12}$, and $S_{64}$ of *Bo* having the same recognition specificities as that of $S_{46}$, $S_8$, $S_{36}$, $S_{47}$, and $S_{41}$ of *Br*, respectively[57, 58]. The pairs exhibited high sequence similarity between eSRK and SP11 alleles, and superposition of the predicted models showed that the mutated residues in each pair were not located around the eSRK–SP11 binding interfaces (**Figure S8**). Conversely, predicted complex models can be used to estimate whether an interspecific *S*-haplotype pair with high sequence similarity is incompatible. In contrast, the pLDDT and pDockQ metrics of the predicted complex models 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. Interestingly, the results suggested that some *S* haplotypes in *Arabidopsis* may have a complex binding mode completely different from that of the *Br* $S_8$ and $S_9$ crystal structures, as shown in **Figure 5A**. MM–GBSA analysis for the predicted complex models of *Ah* $S_{28}$, *Al* $S_6$, $S_{14}$, and $S_a$ showed large negative free energy changes ($\Delta G_{bind}$ < –60 kcal/mol), indicating that these models are plausible (**Figure S9**). The validity of this binding mode needs to be confirmed in future experiments.

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. Our method is also expected to be applicable to other cysteine-rich protein domains that exhibit distinct disulfide classes such as in plant, fungal, and vertebrate α-/β-/θ-defensins[46]. We hope that these computational methodologies will provide insights into other molecular mechanisms involving cysteine-rich protein families.

## METHODS

### Preparation of amino acid sequences

The paired amino acid sequences of eSRK and SP11 from seven species, *Brassica rapa* (*Br*), *Brassica oleracea* (*Bo*), *B. napus* (*Bn*), *R. sativus* (*Rs*), *R. raphanistrum* (*Rr*), *A. lyrata* (*Al*), and *A. halleri* (*Ah*), were obtained from GenBank 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, 29-31, 58-81]. The sequences and accession IDs are shown in **Supplementary File 1**.

### Structure prediction with ColabFold

The structures of eSRK and SP11 proteins were predicted using ColabFold[42], a derivative of AlphaFold2. During the typical prediction process, ColabFold first retrieved an MSA file of the input amino acid sequence using the MMSeqs2 webserver[43, 44]. In this study, we obtained MSA files using the UniRef30[82] (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 through LocalColabFold, the command line interface of ColabFold (https://github.com/YoshitakaMo/localcolabfold).

### Construction of custom MSAs for ColabFold

In addition to the "default" MSA, four other MSAs were constructed to improve the predicted models. The CysBar[49] Python script was used to construct MSAs that properly expressed the structural features of the defensin-like domain of SP11. This tool facilitates the alignment of structurally homologous cysteine residues within MSA. Typically, such homologous Cys residues are identified by combining them with known structures of high homology detected using the DALI server[83, 84]. 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 a 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[35], and 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[50, 51] 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 **Figure S3**.

Additionally, two structure-based MSAs, "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 SP11 $S_8$ and $S_9$ crystal structures 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[35], were subjected to Mustang 3.2.3[85] 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 were most closely related according to the phylogenetic tree (**Figure S5**).

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[85].

Each of the four MSA files comprised 98 sequences. To run the structure prediction using LocalColabFold with the MSA file, we moved the corresponding sequence to the top of the file and reformat it into the a3m format using the "reformat.pl" script bundled in HH-suite[41]. No templates were used for predictions using the four MSAs.

### 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.

### Evaluation of predicted models

We used the pLDDT confidence measure[37], a predicted per-residue lDDT-Cα[86], to evaluate the predicted monomeric model of eSRK or SP11. We also used the two metrics pTM-score[37] and pDockQ[47] 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 by using "pDockQ2.py" available at https://gitlab.com/ElofssonLab/huintaf2/-/blob/main/bin/pDockQ2.py and averaged for the chains in the model.

### 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 eSRK sequences created using Clustal Omega[50,51] 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 **Figure S10**.

### MM–GBSA analysis

MM–GBSA[48] implemented in AmberTools 22[87] was used to calculate the binding free energy, $\Delta G_{\mathrm{bind}}$, 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[88] was used for the proteins, and the system was solvated using the OPC water model[89]. Energy minimization, equilibration, and a 50-ns conventional production run were performed using AMBER 22[87] as per the procedure described in our previous paper[35]. The GBOBC implicit solvent model (parameters $\alpha = 1.0$, $\beta = 0.8$, and $\gamma = 4.85$)[90] 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).

## AUTHOR INFORMATION

### Corresponding Author

moriwaki@bi.a.u-tokyo.ac.jp

### Author Contributions

Y.M, K.M, S.T., K.S., and T.T. conceived this project. T.S., Y.M., and H.J. performed the structure predictions. Y.M. and T.S. wrote this manuscript, which was edited by all other authors. All authors have approved the final version of the manuscript.

### Funding Sources

This research was partially supported by 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.

### Notes

The authors have no competing financial interests to declare.

## ACKNOWLEDGMENT

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. [37]

## ABBREVIATIONS

SRK, *S* receptor kinase; SP11, *S*-locus protein 11; SI, self-incompatibility; pLDDT, predicted local distance difference test; MSA, multiple sequence alignment.

## DATA AVAILABILITY

The MSA files in a3m format and predicted eSRK, SP11, and eSRK–SP11 models using the MSAs in pdb format are freely available at Zenodo, https://doi.org/10.5281/zenodo.8047768.

## REFERENCES

(1) de Nettancourt, D. (2001) *Incompatibility and incongruity in wild and cultivated plants*, 2 ed., Springer, Berlin.

(2) Takayama, S., and Isogai, A. (2005) Self-incompatibility in plants, *Annu. Rev. Plant Biol. 56*, 467–489.

(3) Bateman, A. J. (1955) Self-incompatibility systems in angiosperms: III. Cruciferae, *Heredity 9*, 53–68.

(4) Nou, S., Watanabe, M., Isogai, A., and Hinata, K. (1993) Comparison of *S*-alleles and *S*-glyeoproteins between two wild populations of *Brassica campestris* in Turkey and Japan., *Sexual Plant Reproduction 6*, 79–86.

(5) Ruffio-Chable, V., and Gaude, T. (2001) *S*-haplotype polymorphism in *Brassica oleracea*, *Acta Hortic*, 257–261.

(6) Kao, T. H., and McCubbin, A. G. (1996) How flowering plants discriminate between self and non-self pollen to prevent inbreeding, *Proc. Natl. Acad. Sci. USA 93*, 12059–12065.

(7) Kao, T. H., and Tsukamoto, T. (2004) The molecular and genetic bases of S-RNase-based self-incompatibility, *Plant Cell 16*, S72–S83.

(8) McClure, B. A., Haring, V., Ebert, P. R., Anderson, M. A., Simpson, R. J., Sakiyama, F., and Clarke, A. E. (1989) Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases, *Nature 342*, 955–957.

(9) Franklin-Tong, N., and Franklin, F. C. H. (2003) Gametophytic self-incompatibility inhibits pollen tube growth using different mechanisms, *Trends Plant Sci. 8*, 598–605.

(10) Foote, H. C. C., Ride, J. P., Franklintong, V. E., Walker, E. A., Lawrence, M. J., and Franklin, F. C. H. (1994) Cloning and expression of a distinctive class of self-incompatibility (*S*) gene from *Papaver rhoeas* L, *Proc. Natl. Acad. Sci. USA 91*, 2265–2269.

(11) Lane, M. D., and Lawrence, M. J. (1993) The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. VII. The number of *S*-alleles in the species, *Heredity 71*, 596–602.

(12) Hearn, M. J., Franklin, F. C. H., and Ride, J. P. (1996) Identification of a membrane glycoprotein in pollen of Papaver rhoeas which binds stigmatic self-incompatibility (S-) proteins, *Plant J. 9*, 467–475.

(13) Jordan, N. D., Kakeda, K., Conner, A., Ride, J. P., Franklin-Tong, V. E., and Franklin, F. C. H. (1999) S-protein mutants indicate a functional role for SBP in the self-incompatibility reaction of *Papaver rhoeas*, *Plant J. 20*, 119–125.

(14) Nasrallah, M. E., and Wallace, D. H. (1967) Immunogenetics of self-incompatibility in *Brassica oleracea L.*, *Heredity 22*, 519–527.

(15) Nishio, T., and Hinata, K. (1977) Analysis of S-specific proteins in stigma of *Brassica oleracea* L. by isoelectric focusing, *Heredity 38*, 391–396.

(16) Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A., and Hinata, K. (2000) The *S* receptor kinase determines self-incompatibility in *Brassica* stigma, *Nature 403*, 913–916.

(17) Stein, J. C., Howlett, B., Boyes, D. C., Nasrallah, M. E., and Nasrallah, J. B. (1991) Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*, *Proc. Natl. Acad. Sci. USA 88*, 8816–8820.

(18) Takayama, S., Shiba, H., Iwano, M., Shimosato, H., Che, F. S., Kai, N., Watanabe, M., Suzuki, G., Hinata, K., and Isogai, A. (2000) The pollen determinant of self-incompatibility in *Brassica campestris*, *Proc. Natl. Acad. Sci. USA 97*, 1920–1925.

(19) Schopfer, C. R., Nasrallah, M. E., and Nasrallah, J. B. (1999) The male determinant of self-incompatibility in *Brassica*, *Science 286*, 1697–1700.

(20) Shiba, H., Takayama, S., Iwano, M., Shimosato, H., Funato, M., Nakagawa, T., Che, F. S., Suzuki, G., Watanabe, M., Hinata, K., and Isogai, A. (2001) A pollen coat protein, SP11/SCR, determines the pollen *S*-specificity in the self-incompatibility of *Brassica* species, *Plant Physiol. 125*, 2095–2103.

(21) Suzuki, G., Kai, N., Hirose, T., Fukui, K., Nishio, T., Takayama, S., Isogai, A., Watanabe, M., and Hinata, K. (1999) 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 153*, 391–400.

(22) Takayama, S., Shimosato, H., Shiba, H., Funato, M., Che, F. S., Watanabe, M., Iwano, M., and Isogai, A. (2001) Direct ligand-receptor complex interaction controls *Brassica* self-incompatibility, *Nature 413*, 534–538.

(23) Kachroo, A., Schopfer, C. R., Nasrallah, M. E., and Nasrallah, J. B. (2001) Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility, *Science 293*, 1824–1826.

(24) Shimosato, H., Yokota, N., Shiba, H., Iwano, M., Entani, T., Che, F. S., Watanabe, M., Isogai, A., and Takayama, S. (2007) Characterization of the SP11/SCR high-affinity binding site involved in self/nonself recognition in *Brassica* self-incompatibility, *Plant Cell 19*, 107–117.

(25) Cabrillac, D., Cock, J. M., Dumas, C., and Gaude, T. (2001) The S-locus receptor kinase is inhibited by thioredoxins and activated by pollen coat proteins, *Nature 410*, 220–223.

(26) Boyes, D. C., and Nasrallah, J. B. (1993) Physical linkage of the *SLG* and *SRK* genes at the self-incompatibility locus of *Brassica oleracea*, *Mol. Gen. Genet. 236*, 369–373.

(27) Naithani, S., Chookajorn, T., Ripoll, D. R., and Nasrallah, J. B. (2007) Structural modules for receptor dimerization in the *S*-locus receptor kinase extracellular domain, *Proc. Natl. Acad. Sci. USA 104*, 12211–12216.

(28) Kusaba, M., Nishio, T., Satta, Y., Hinata, K., and Ockendon, D. (1997) 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 94*, 7673–7678.

(29) Sato, K., Nishio, T., Kimura, R., Kusaba, M., Suzuki, T., Hatakeyama, K., Ockendon, D. J., and Satta, Y. (2002) Coevolution of the *S*-locus genes *SRK*, *SLG* and *SP11/SCR* in *Brassica oleracea* and *B. rapa*, *Genetics 162*, 931–940.

(30) Watanabe, M., Ito, A., Takada, Y., Ninomiya, C., Kakizaki, T., Takahata, Y., Hatakeyama, K., Hinata, K., Suzuki, G., Takasaki, T., Satta, Y., Shiba, H., Takayama, S., and Isogai, A. (2000) Highly divergent sequences of the pollen self-incompatibility (*S*) gene in class-I *S* haplotypes of *Brassica campestris* (syn. *rapa*) L, *FEBS Lett. 473*, 139–144.

(31) Shiba, H., Iwano, M., Entani, T., Ishimoto, K., Shimosato, H., Che, F. S., Satta, Y., Ito, A., Takada, Y., Watanabe, M., Isogai, A., and Takayama, S. (2002) The dominance of alleles controlling self-incompatibility in *Brassica* pollen is regulated at the RNA level, *Plant Cell 14*, 491–504.

(32) Tarutani, Y., Shiba, H., Iwano, M., Kakizaki, T., Suzuki, G., Watanabe, M., Isogai, A., and Takayama, S. (2010) Trans-acting small RNA determines dominance relationships in *Brassica* self-incompatibility, *Nature 466*, 983–986.

(33) Yasuda, S., Wada, Y., Kakizaki, T., Tarutani, Y., Miura-Uno, E., Murase, K., Fujii, S., Hioki, T., Shimoda, T., Takada, Y., Shiba, H., Takasaki-Yasuda, T., Suzuki, G., Watanabe, M., and Takayama, S. (2017) A complex dominance hierarchy is controlled by polymorphism of small RNAs and their targets, *Nat. Plants 3*.

(34) Ma, R., Han, Z. F., Hu, Z. H., Lin, G. Z., Gong, X. Q., Zhang, H. Q., Nasrallah, J. B., and Chai, J. J. (2016) Structural basis for specific self-incompatibility response in *Brassica*, *Cell Res. 26*, 1320–1329.

(35) Murase, K., Moriwaki, Y., Mori, T., Liu, X., Masaka, C., Takada, Y., Maesaki, R., Mishima, M., Fujii, S., Hirano, Y., Kawabe, Z., Nagata, K., Terada, T., Suzuki, G., Watanabe, M., Shimizu, K., Hakoshima, T., and Takayama, S. (2020) Mechanism of self/nonself-discrimination in *Brassica* self-incompatibility, *Nat. Commun. 11*.

(36) Ockendon, D. J. (2000) The *S*-allele collection of *Brassica oleracea*, *Acta Hortic. 539*, 25–30.

(37) Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold, *Nature 596*, 583–589.

(38) Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., and Hassabis, D. (2022) Protein complex prediction with AlphaFold-Multimer, *bioRxiv*, 2021 2010 2004.463034.

(39) Johnson, L. S., Eddy, S. R., and Portugaly, E. (2010) Hidden Markov model speed heuristic and iterative HMM search procedure, *BMC Bioinformatics 11*.

(40) Remmert, M., Biegert, A., Hauser, A., and Soding, J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment, *Nat. Methods 9*, 173–175.

(41) Steinegger, M., Meier, M., Mirdita, M., Vohringer, H., Haunsberger, S. J., and Soding, J. (2019) HH-suite3 for fast remote homology detection and deep protein annotation, *BMC Bioinformatics 20*, 473.

(42) Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all, *Nat. Methods 19*, 679–682.

(43) Steinegger, M., and Soding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets, *Nat. Biotechnol. 35*, 1026–1028.

(44) Mirdita, M., Steinegger, M., and Soding, J. (2019) MMseqs2 desktop and local web server app for fast, interactive sequence searches, *Bioinformatics 35*, 2856-2858.

(45) Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zidek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., and Velankar, S. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, *Nucleic Acids Res. 50*, D439–D444.

(46) Shafee, T. M. A., Lay, F. T., Hulett, M. D., and Anderson, M. A. (2016) The defensins consist of two independent, convergent protein superfamilies, *Mol. Biol. Evol. 33*, 2345–2356.

(47) Bryant, P., Pozzati, G., and Elofsson, A. (2022) Improved prediction of protein-protein interactions using AlphaFold2, *Nat. Commun. 13*.

(48) Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S. H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models, *Acc. Chem. Res. 33*, 889–897.

(49) Shafee, T. M. A., Robinson, A. J., van der Weerden, N., and Anderson, M. A. (2016) Structural homology guided alignment of cysteine rich proteins, *Springerplus 5*.

(50) Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W. Z., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol. 7*.

(51) Sievers, F., and Higgins, D. G. (2018) Clustal Omega for making accurate alignments of many protein sequences, *Protein Sci. 27*, 135–145.

(52) Weigt, M., White, R. A., Szurmant, H., Hoch, J. A., and Hwa, T. (2009) Identification of direct residue contacts in protein-protein interaction by message passing, *Proc. Natl. Acad. Sci. USA 106*, 67–72.

(53) Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., Zecchina, R., Onuchic, J. N., Hwa, T., and Weigt, M. (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families, *Proc. Natl. Acad. Sci. USA 108*, E1293–E1301.

(54) Balakrishnan, S., Kamisetty, H., Carbonell, J. G., Lee, S. I., and Langmead, C. J. (2011) Learning generative models for protein fold families, *Proteins: Struct. Funct. Bioinform. 79*, 1061–1078.

(55) Ekeberg, M., Lovkvist, C., Lan, Y. H., Weigt, M., and Aurell, E. (2013) Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, 87*, 012707.

(56) Kamisetty, H., Ovchinnikov, S., and Baker, D. (2013) Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era, *Proc. Natl. Acad. Sci. USA 110*, 15674–15679.

(57) Kimura, R., Sato, K., Fujimoto, R., and Nishio, T. (2002) Recognition specificity of self-incompatibility maintained after the divergence of *Brassica oleracea* and *Brassica rapa*, *Plant J. 29*, 215–223.

(58) Sato, Y., Fujimoto, R., Toriyama, K., and Nishio, T. (2003) Commonality of self-recognition specificity of *S* haplotypes between *Brassica oleracea* and *Brassica rapa*, *Plant Mol. Biol. 52*, 617–626.

(59) Glavin, T. L., Goring, D. R., Schafer, U., and Rothstein, S. J. (1994) Features of the extracellular domain of the *S*-locus receptor kinase from *Brassica*, *Mol. Gen. Genet. 244*, 630–637.

(60) Fukai, E., Fujimoto, R., and Nishio, T. (2003) Genomic organization of the *S* core region and the *S* flanking regions of a class-II *S* haplotype in *Brassica rapa*, *Mol. Genet. Genomics 269*, 361–369.

(61) Vanoosthuyse, V., Miege, C., Dumas, C., and Cock, J. M. (2001) 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. 46*, 17–34.

(62) Boggs, N. A., Dwyer, K. G., Shah, P., McCuuoch, A. A., Bechsgaard, J., Schierup, M. H., Nasrallah, M. E., and Nasrallah, J. B. (2009) Expression of distinct self-incompatibility specificities in *Arabidopsis thaliana*, *Genetics 182*, 1313–1321.

(63) Delorme, V., Giranton, J. L., Hatzfeld, Y., Friry, A., Heizmann, P., Ariza, M. J., Dumas, C., Gaude, T., and Cock, J. M. (1995) 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. 7*, 429–440.

(64) Suzuki, T., Kusaba, M., Matsushita, M., Okazaki, K., and Nishio, T. (2000) Characterization of *Brassica S*-haplotypes lacking *S*-locus glycoprotein, *FEBS Lett. 482*, 102–108.

(65) Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J. B., and Nasrallah, M. E. (2001) Self-incompatibility in the genus *Arabidopsis*: Characterization of the *S* locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*, *Plant Cell 13*, 627–643.

(66) Takuno, S., Oikawa, E., Kitashiba, H., and Nishio, T. (2010) Assessment of genetic diversity of accessions in Brassicaceae genetic resources by frequency distribution analysis of *S* haplotypes, *Theor. Appl. Genet. 120*, 1129–1138.

(67) Kusaba, M., Matsushita, M., Okazaki, K., Satta, Y., and Nishio, T. (2000) Sequence and structural diversity of the *S* locus genes from different lines with the same self-recognition specificities in *Brassica oleracea*, *Genetics 154*, 413–420.

(68) Yamakawa, S., Watanabe, M., Hinata, K., Suzuki, A., and Isogai, A. (1995) The Sequences of *S*-Receptor Kinases (SRK) Involved in Self-incompatibility and Their Homologies to *S*-Locus Glycoproteins of *Brassica campestris*, *Biosci., Biotechnol., Biochem. 59*, 161–162.

(69) Kumar, V., and Trick, M. (1994) Expression of the *S*-locus receptor kinase multigene family in *Brassica oleracea*, *Plant J. 6*, 807–813.

(70) Goubet, P. M., Berges, H., Bellec, A., Prat, E., Helmstetter, N., Mangenot, S., Gallina, S., Holl, A. C., Fobis-Loisy, I., Vekemans, X., and Castric, V. (2012) Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in *Arabidopsis*, *PLoS Genet. 8*, e1002495.

(71) Kakizaki, T., Takada, Y., Fujioka, T., Suzuki, G., Satta, Y., Shiba, H., Isogai, A., Takayama, S., and Watanabe, M. (2006) Comparative analysis of the *S*-intergenic region in class-II *S* haplotypes of self-incompatible *Brassica rapa* (syn. *campestris*), *Genes Genet. Syst. 81*, 63–67.

(72) Cabrillac, D., Delorme, V., Garin, J., Ruffio-Chable, V., Giranton, J. L., Dumas, C., Gaude, T., and Cock, J. M. (1999) The $S_{15}$ self-Incompatibility haplotype in *Brassica oleracea* includes three *S* gene family members expressed in stigmas, *Plant Cell 11*, 971–986.

(73) Bechsgaard, J. S., Castric, V., Vekemans, X., and Schierup, M. H. (2006) The transition to self-compatibility in *Arabidopsis thaliana* and evolution within *S*-haplotypes over 10 Myr, *Mol. Biol. Evol. 23*, 1741–1750.

(74) Fujimoto, R., Okazaki, K., Fukai, E., Kusaba, M., and Nishio, T. (2006) Comparison of the genome structure of the self-incompatibility (*S*) locus in interspecific pairs of *S* haplotypes, *Genetics 173*, 1157–1167.

(75) Dwyer, K. G., Berger, M. T., Ahmed, R., Hritzo, M. K., McCulloch, A. A., Price, M. J., Serniak, N. J., Walsh, L. T., Nasrallah, J. B., and Nasrallah, M. E. (2013) Molecular characterization and evolution of self-Incompatibility genes in *Arabidopsis thaliana*: The case of the *Sc* haplotype, *Genetics 193*, 985–994.

(76) Kusaba, M., and Nishio, T. (1999) Comparative analysis of *S* haplotypes with very similar *SLG* alleles in *Brassica rapa* and *Brassica oleracea*, *Plant J. 17*, 83–91.

(77) Hatakeyama, K., Takasaki, T., Watanabe, M., and Hinata, K. (1998) Molecular characterization of *S* locus genes, *SLG* and *SRK*, in a pollen-recessive self-incompatibility haplotype of *Brassica rapa* L, *Genetics 149*, 1587–1597.

(78) Okamoto, S., Odashima, M., Fujimoto, R., Sato, Y., Kitashiba, H., and Nishio, T. (2007) Self-compatibility in *Brassica napus* is caused by independent mutations in *S*-locus genes, *Plant J. 50*, 391–400.

(79) Suzuki, G., Watanabe, M., Isogai, A., and Hinata, K. (1997) Highly conserved 5'-flanking regions of two self-incompatibility genes, $SLG^9$ and $SRK^9$, *Gene 191*, 123–126.

(80) Takuno, S., Fujimoto, R., Sugimura, T., Sato, K., Okamoto, S., Zhang, S. L., and Nishio, T. (2007) Effects of recombination on hitchhiking diversity in the *Brassica* self-incompatibility locus complex, *Genetics 177*, 949–958.

(81) Boggs, N. A., Nasrallah, J. B., and Nasrallah, M. E. (2009) Independent *S*-Locus mutations caused self-fertility in *Arabidopsis thaliana*, *PLoS Genet. 5*, e1000426.

(82) Mirdita, M., von den Driesch, L., Galiez, C., Martin, M. J., Soding, J., and Steinegger, M. (2017) Uniclust databases of clustered and deeply annotated protein sequences and alignments, *Nucleic Acids Res. 45*, D170–D176.

(83) Holm, L. (2022) Dali server: structural unification of protein families, *Nucleic Acids Res. 50*, W210–W215.

(84) Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res. 38*, W545–W549.

(85) Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J., and Lesk, A. M. (2006) MUSTANG: A multiple structural alignment algorithm, *Proteins: Struct. Funct. Bioinform. 64*, 559–574.

(86) Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013) lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests, *Bioinformatics 29*, 2722-2728.

(87) Case, D. A., Aktulga, H. M., Belfon, K., Ben-Shalom, I. Y., Berryman, J. T., Brozell, S. R., Cerutti, D. S., Cheatham, T. E. I., Cisneros, G. A., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Giambasu, G., Gilson, M. K., Gohlke, H., Goetz, A. W., Harris, R., Izadi, S., Izmailov, S. A., Kasavajhala, K., Kaymak, M. C., King, E., Kovalenko, A., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Machado, M., Man, V., Manathunga, M., Merz, K. M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., O'Hearn, K. A., Onufriev, A., Pan, F., Pantano, S., Qi, R., Rahnamoun, A., Roe, D. R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shajan, A., Shen, J., Simmerling, C. L., Skrynnikov, N. R., Smith, J., Swails, J., Walker, R. C., Wang, J., Wang, J., Wei, H., Wolf, R. M., Wu, X., Xiong, Y., Xue, Y., York, D. M., Zhao, S., and P.A., K. (2022) AMBER 22, University of California, San Francisco.

(88) Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., Bickel, J., Wang, Y. Z., Pincay, J., Wu, Q., and Simmerling, C. (2020) ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energies in solution, *J. Chem. Theory Comput. 16*, 528–552.

(89) Izadi, S., Anandakrishnan, R., and Onufriev, A. V. (2014) Building Water Models: A Different Approach, *Journal of Physical Chemistry Letters 5*, 3863–3871.

(90) Roe, D. R., Okur, A., Wickstrom, L., Hornak, V., and Simmerling, C. (2007) 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 111*, 1846–1857.

(91) Wu, T. Q., Hou, J., Adhikari, B., and Cheng, J. L. (2020) Analysis of several key factors influencing deep learning-based inter-residue contact prediction, *Bioinformatics 36*, 1091–1098.

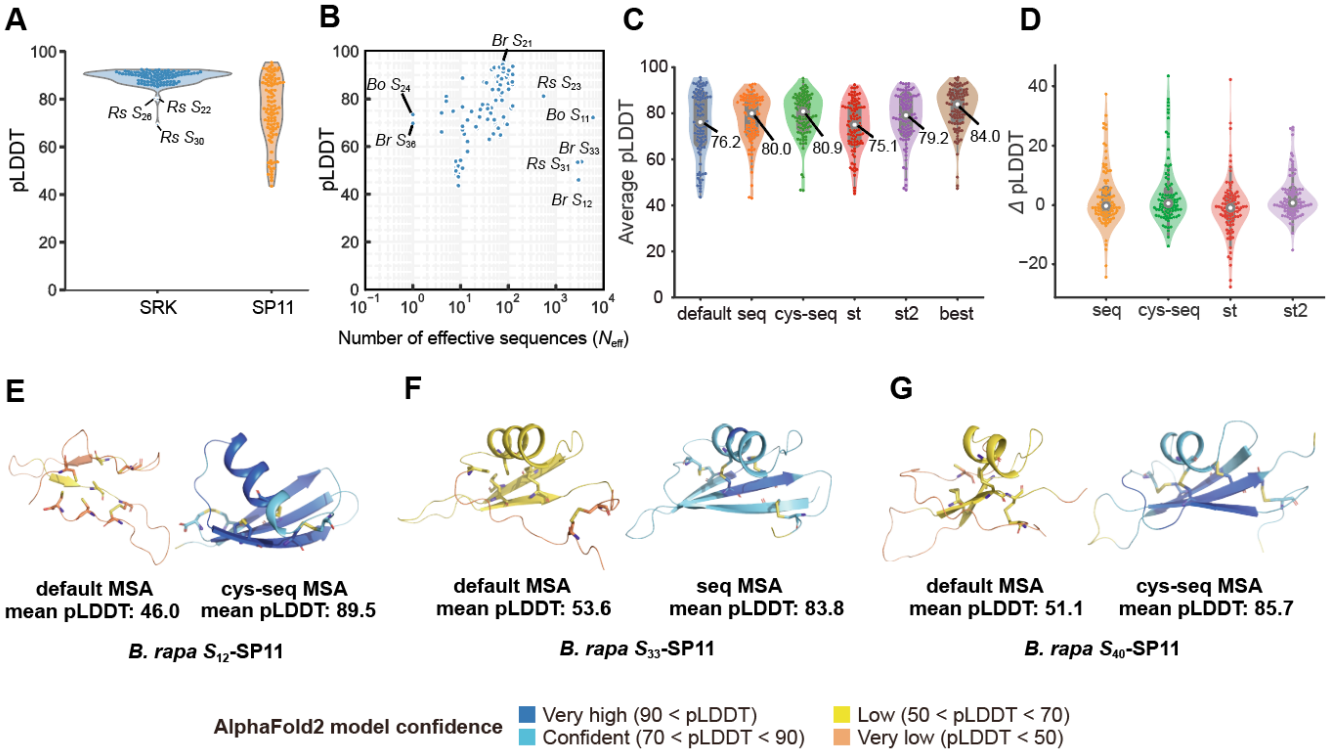



**Figure 1. Structure prediction of S-locus protein 11 (SP11) proteins with ColabFold using various multiple sequence alignments (MSAs).** (A) The mean predicted local distance difference test (pLDDT) of *S* receptor kinase (SRK) and SP11 models predicted using ColabFold with the "default_MSA" (uniref30_2202 + 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[91] 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.

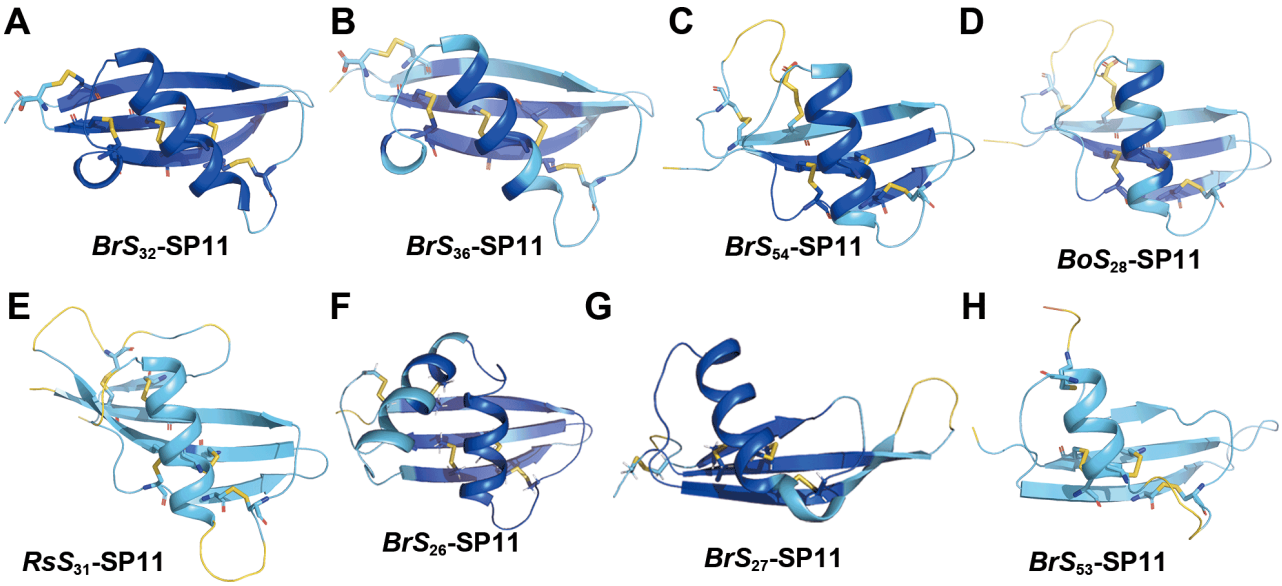



**Figure 2. Predicted novel class I SP11 structures and disulfide bond connectivity.** (A–H) *Br* $S_{32}$ (A), *Br* $S_{36}$ (B), *Br* $S_{54}$ (C), *Bo* $S_{28}$ (D), *Rs* $S_{31}$ (E), *Br* $S_{26}$ (F), *Br* $S_{27}$ (G), *Br* $S_{53}$ (H)-SP11 model structures predicted using ColabFold with custom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

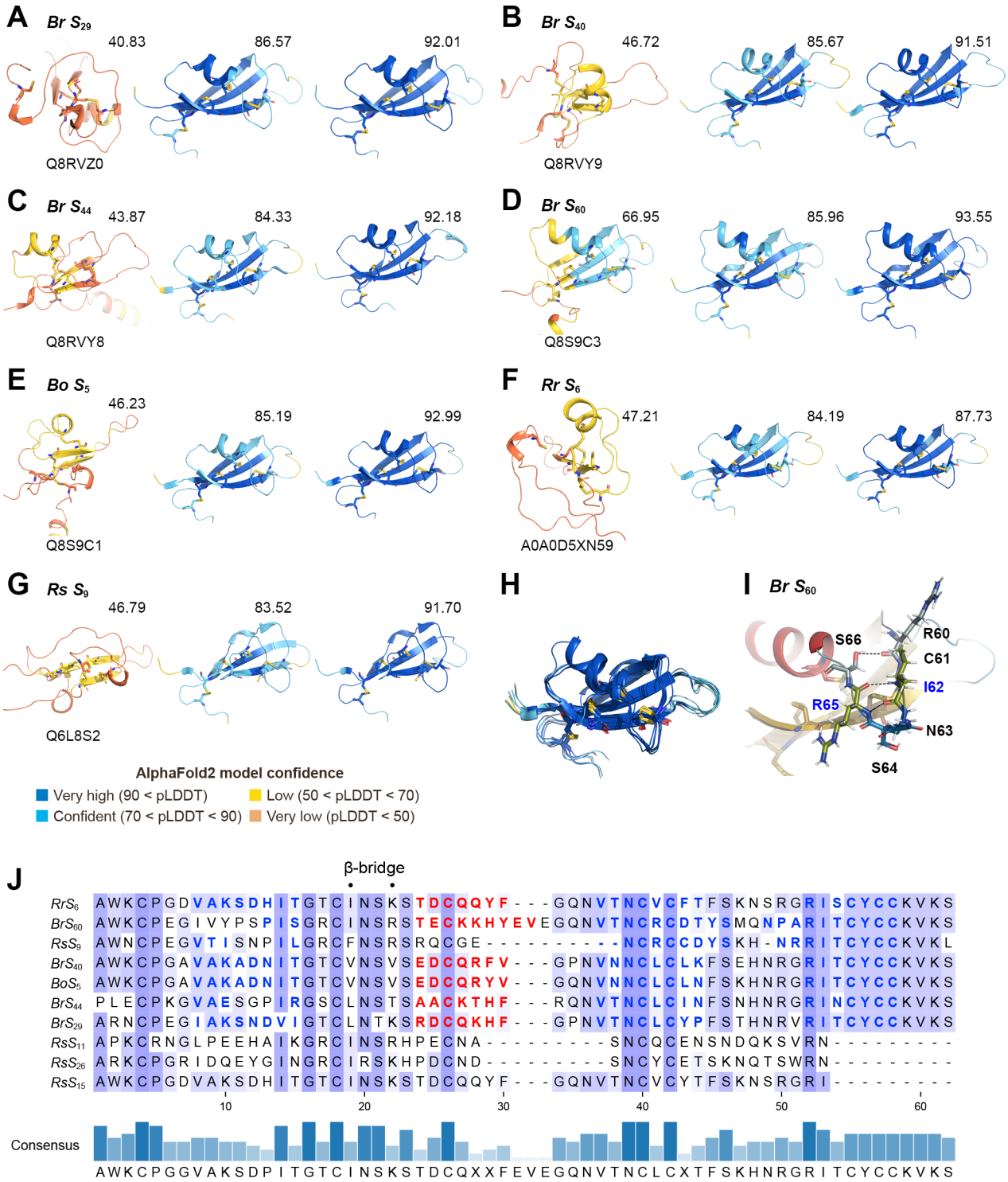



**Figure 3. Predicted models of class-II SP11s.** (A–G) Predicted SP11 models of *Br* $S_{29}$ (A), *Br* $S_{40}$ (B), *Br* $S_{44}$ (C), *Br* $S_{60}$ (D), *Bo* $S_5$ (E), *Rr* $S_6$ (F), and *Rs* $S_9$ (G) haplotypes. The model deposited in AlphaFold DB (version 2022-11-01), the best one among the 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 by "·" marks.

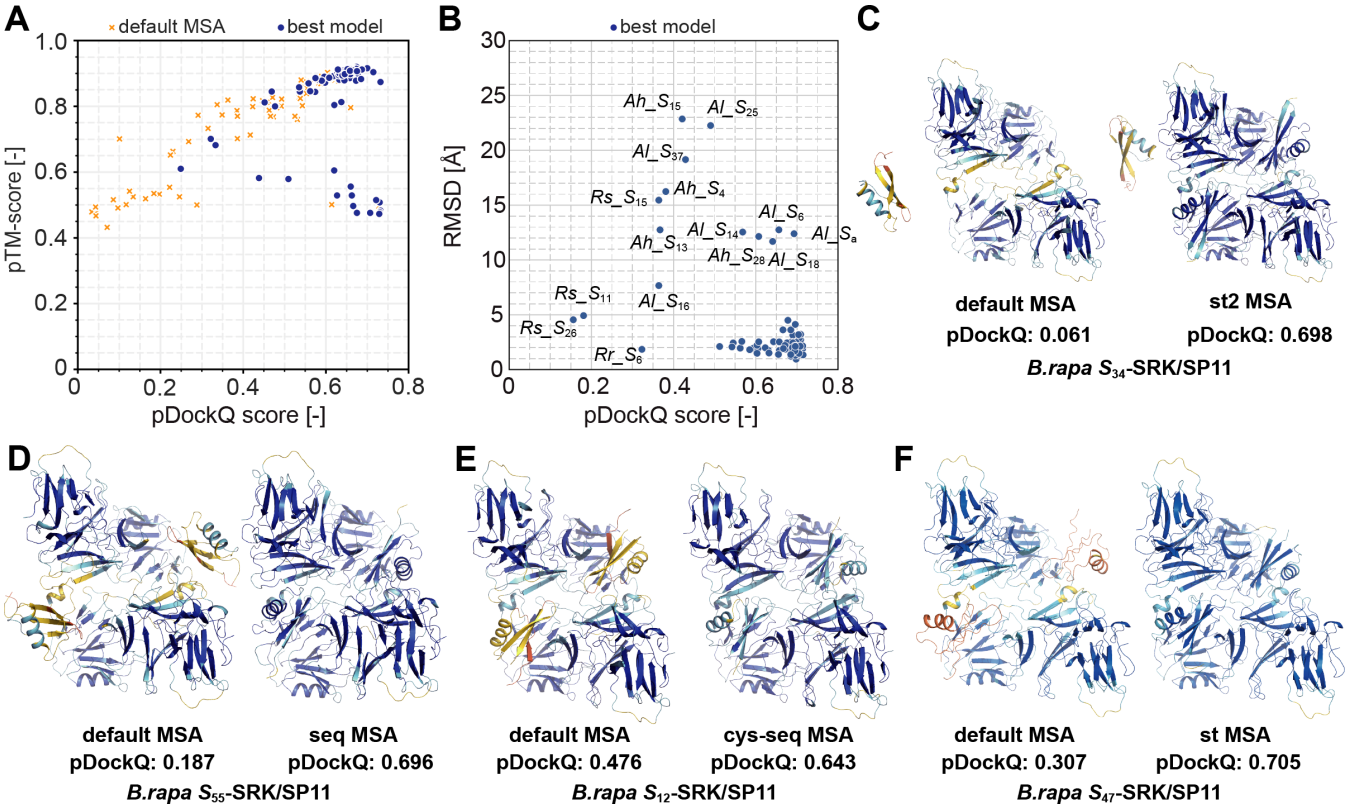



**Figure 4. Complex predictions with the custom MSAs.** (A) Correlation between pTM-score and pDockQ for the predicted complex models. Orange crosses and blue circles indicate the complex models predicted using the default MSA and the highest pDockQ of the four MSAs, respectively. (B) Root mean square 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.

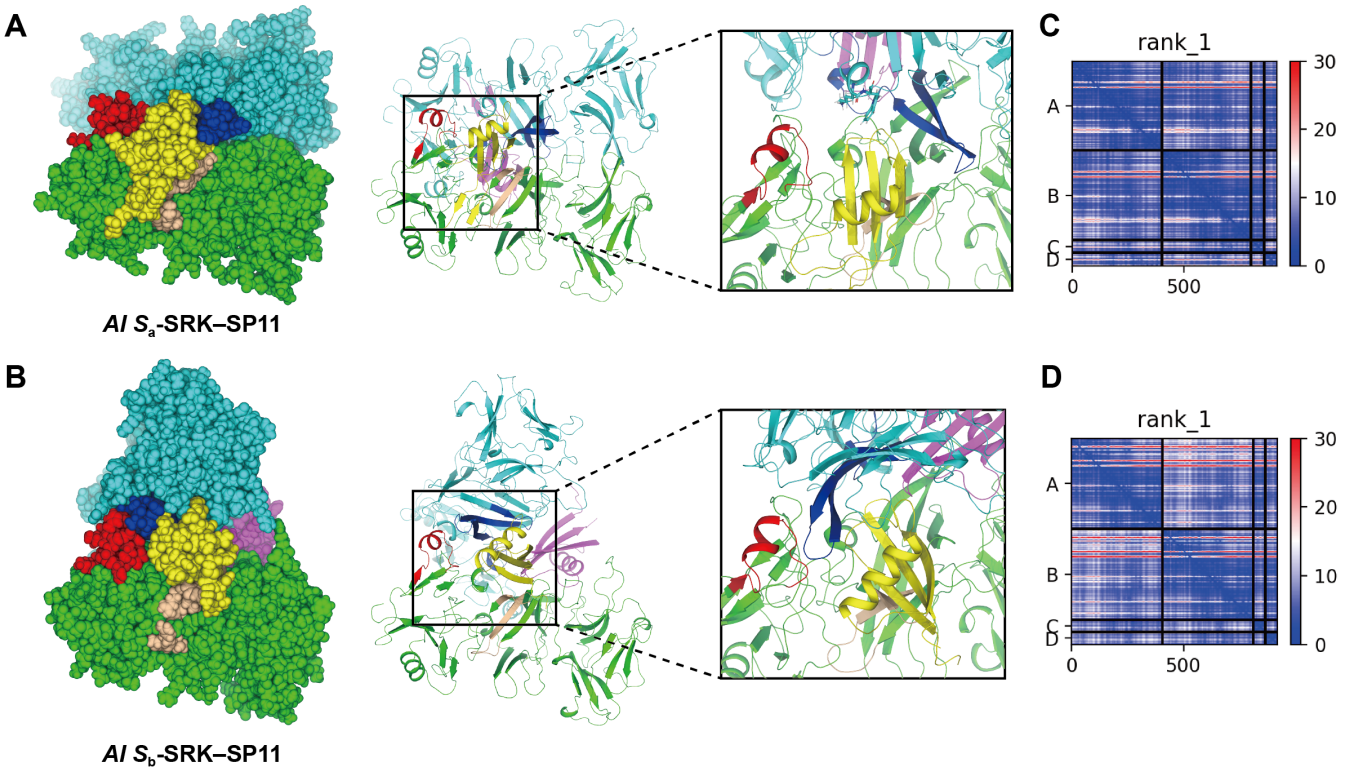



**Figure 5. Predicted binding mode for some *Al* or *Ah* SRK–SP11 complexes.** (A–B) Predicted *Al* $S_a$ (A) and *Al* $S_b$ (B) complex models using ColabFold. Two SRK monomers are colored in green and cyan, and two SP11s are in yellow and purple. The position and orientation of *Al* $S_a$- and *Al* $S_b$- SRK chain A (green) are aligned. The HV regions I, II, and III of SRK are colored in red, deep blue, and wheat, 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.

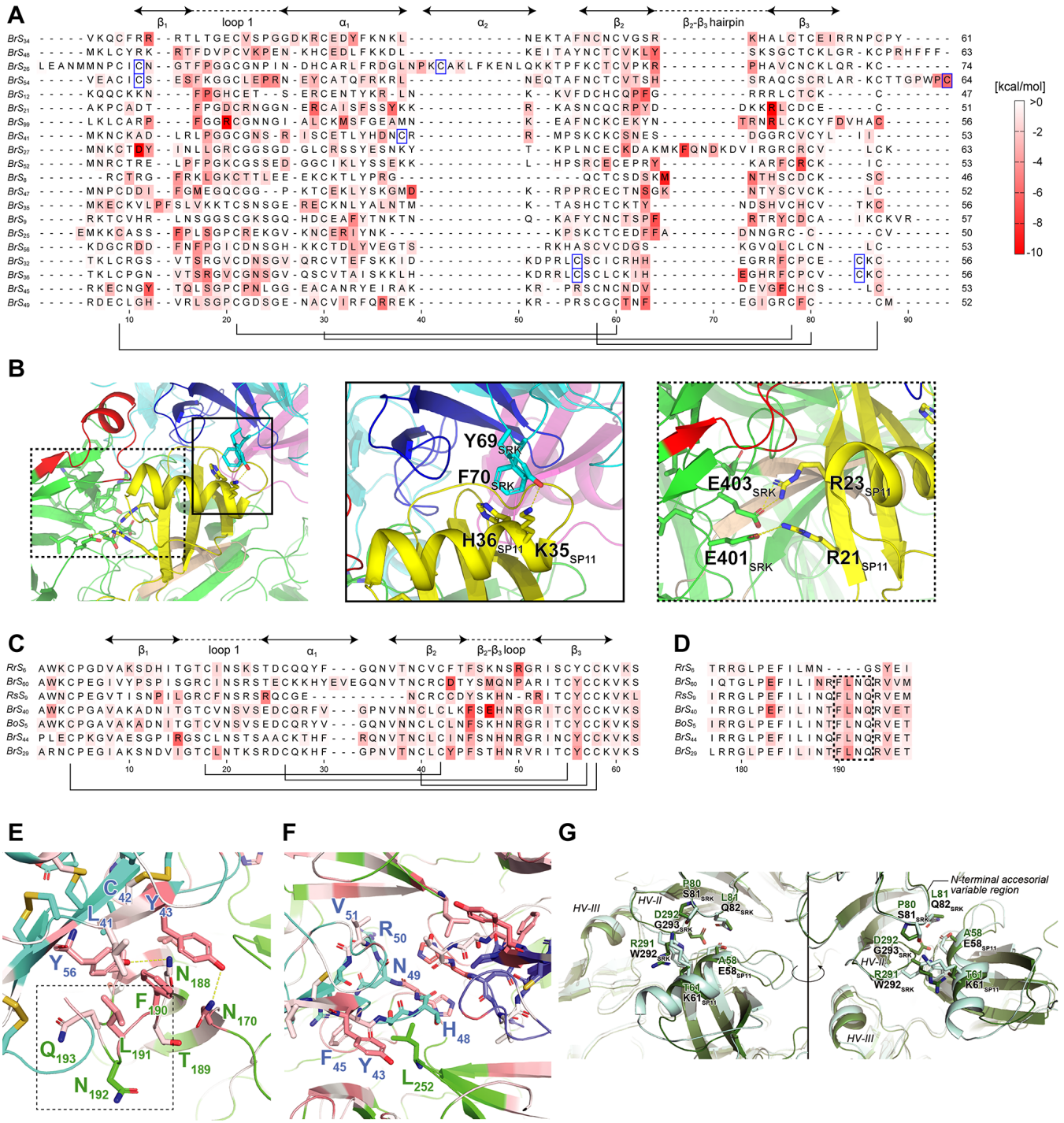



**Figure 6. Molecular mechanics–generalized born surface area (MM–GBSA) analysis using the predicted structures of SRK–SP11 complexes.** (A) *Br* class-I SP11 sequences with heat maps indicating energy contributions from each residue. Cys residue pairs forming disulfide bonds are shown in black lines. 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 view 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 Figure 4. (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 HV-I region of eSRK (E) and the β2-β3 loop of SP11 (F) in the predicted model of *Br* $S_{29}$. eSRK and SP11 are colored as cyan and green, respectively. Residues at the interface are colored according to their per-residue $\Delta G_{\mathbf{bind}}$ 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.

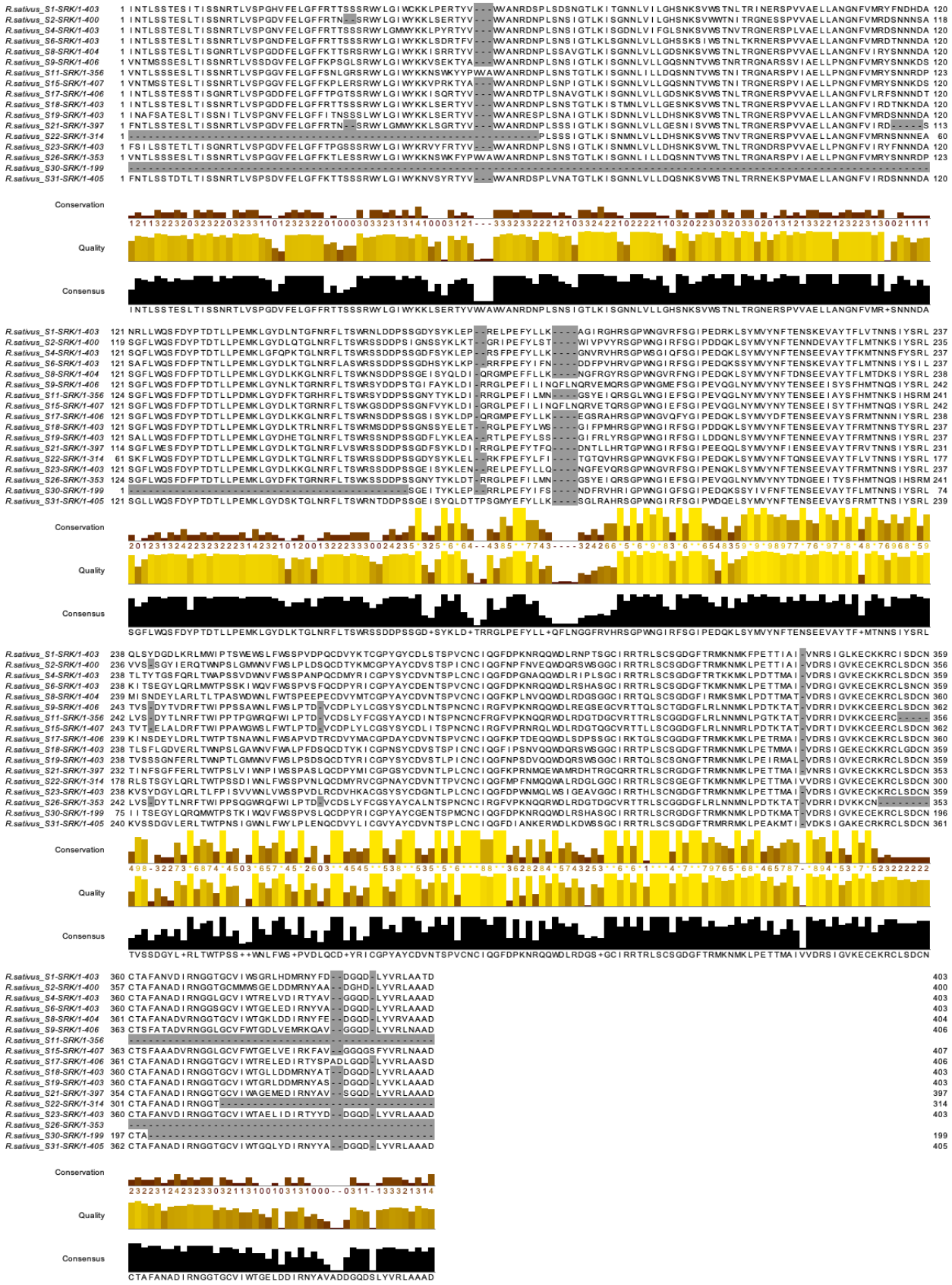



**Figure S1. Sequence alignment of *Rs* SRKs.** Sequence regions used to run the structure prediction are shown. This alignment has been visualized using Jalview 2.11.2.6.

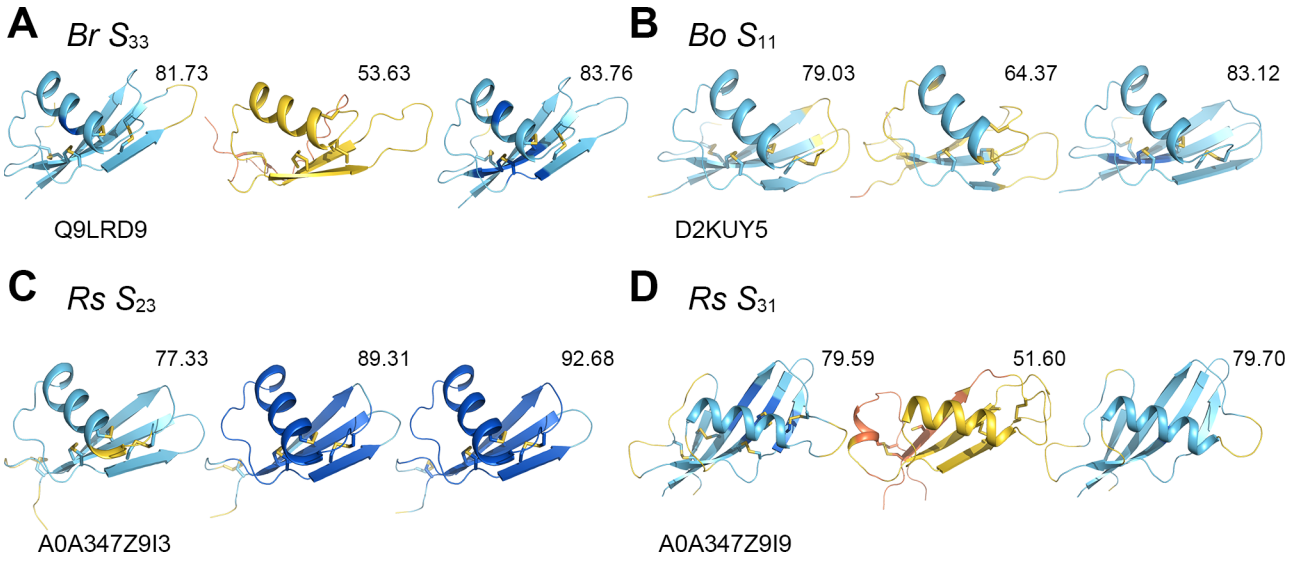



**Figure S2. Improved SP11 models with custom MSAs.** (A–D) Predicted SP11 structures of *Br* $S_{33}$ (A), *Bo* $S_{11}$ (B), *Rs* $S_{23}$ (C), and *Rs* $S_{31}$ (D) haplotypes. The predicted model deposited in AlphaFold DB (version 2022-11-01), one with the default MSA, and the best one among the four custom MSAs are depicted in the left, middle, and right panels, respectively. UniProt accession IDs are shown under the model of AlphaFold DB. The mean per-residue pLDDT value is shown on the upper right corner of each model. The structures are colored according to the AlphaFold pLDDT model confidence.

```
                                  10        20        30        40        50        60        70        80        90       100
SP11-8_rapa/1-46                  ------RCTR----GFRK-LGKC-TTLEE------------E-KCKTLYPR---------GQCTCSDSKM--------NTHSCDCKSC----------
SP11-9_rapa/1-57                  RK----TCVH----RLNS-GGSC-GKSGQ------------H-DCEAFYTNK--T-NQKAFYCNCTSPFR---------TRYCDCAIKCKVR----------
SP11-12_rapa/1-47                 ------KQQC----KKNF-PGHC-ETSE---------------RCENTYKRL----NKKVFDCHCQPFGR---------RRLCTCKC----------
SP11-21_rapa/1-51                 ------AKPC----ADTF-PGDC-RNGGN------------E-RCAISFSSY--K-KRKASNCQCRPYDDKKR--------LCDCEC----------
SP11-22_rapa/1-60                 QK----LCIC----NGTF-EGRC-ANTRD------------V-FCENLFKNN--LQEKTAFNCGCTD---SN------SHILCTCHVVRPAQLC----------
SP11-25_rapa/1-50                 EMK---KCAS----SFPL-SGPC-REKGV------------K-NCERIYNK-------KPSKCTCEDFFAD-------NNGRCCCV----------
SP11-26_rapa/1-51                 ----------------------------NDHCARLFRDGLNP-KCAKLFKEN--LQKKTPFKCTCVP---KR------PHAVCNCKLQRKC----------
SP11-27_rapa/1-63                 MN----KCTD----YINL-LGRC-GGSGD------------G-LCRSSYESN--K-YTKPLNCECKDAKMKFQNDKDVIRGRCRCVLCK----------
SP11-29_rapa/1-59                 AR----NCPEGIAKSNDV-IGTC-LNTKS------------R-DCQKHF-------GPNVTNCLCYPFSTH-------NRVRITCYCCKVKS----------
SP11-32_rapa/1-56                 TK----LCRG----SVTS-RGVC-DNSGV------------Q-RCVTEFSKK--I-DKDPRLCSCICRHHE-------GRRFCPCECKC----------
SP11-33_rapa/1-60                 ------QFLC----QRNF-RGEG-CVS-P------------A-DCETLFKNK--LNEKTASNCTCVRKTGKY------ARAQCTCRLPHKCVPR----------
SP11-34_rapa/1-61                 VK----QCFR----RRTL-TGEC-VSPGGD-----------K-RCEDYFKNK--LNEKTAFNCNCVG---SR------KHALCTCEIRRNPCPY----------
SP11-35_rapa/1-56                 MK----ECKVLP--FSLV-KKTC-SNSGE------------R-ECKNLYALN--TM-KTASHCTCKTYN---------DSHVCHCVTKC----------
SP11-36_rapa/1-56                 TK----LCPG----NVTS-RGVC-GNSGV------------Q-SCVTAISKK--L-HKDRRLCSCLCKIHE-------GHRFCPCVCKC----------
SP11-37_rapa/1-59                 ------LKRC----DKSF-PGDC-GNNGK------------I-VCEKSWNAV----KKKAFNCICDIYILKKD------RRICKCDLDVHICPL----------
SP11-40_rapa/1-59                 AW----KCPGAVAKADNI-TGTC-VNSVS------------E-DCQRFV-------GPNVNNCLCLKFSEH-------NRGRITCYCCKVKS----------
SP11-41_rapa/1-53                 MK----NCKA----DLRL-PGGC-GNSR-------------I-SCETLYHDN--C-RRMPSKCKCSNE--S-------DGGRCVCYLII----------
SP11-44_rapa/1-59                 PL----ECPKGVAESGPI-RGSC-LNSTS------------A-ACKTHF-------RQNVTNCLCINFSNH-------NRGRINCYCCKVKS----------
SP11-45_rapa/1-53                 RK----ECNG----YTQL-SGPC-PNLGG------------E-ACANRYEIR--A-KKRPRSCNCNDVDE--------VGFCHCSLC----------
SP11-46_rapa/1-53                 SDMK--QCKI----GYRM-PGNC-AELA-------------Q-TCENFCSRG--KE-KKPSHWKCTNGPK--------NTYSCDCK----------
SP11-47_rapa/1-52                 MN----PCDD----IFGM-EGQC-GGP--------------K-TCEKLYSKG--MD-KRPPRCECTNSGK--------NTYSCVCKLC----------
SP11-48_rapa/1-63                 MK----LCYR----KRTF-DVPC-VKP-EN-----------K-HCEDLFKKD--LKEITAYNCTCVK---LY------SKSGCTCKLGRKCPRHFFF----------
SP11-49_rapa/1-52                 RD----ECLG----HVRL-SGPC-GDSGE------------N-ACVIRFQRR--E-KKRPRSCGCTNFEG---------IGRCFCCM----------
SP11-52_rapa/1-53                 MN----RCTR----ELPF-PGKC-GSSED------------G-GCIKLYSSE--KK-LHPSRCECEPRY---------KARFCRCKIC----------
SP11-53_rapa/1-57                 LN----NCKD----ALRL-PGLC-YSGT-------------L-TCEGLYHKV--C-GRKAKSCNCFNDFSNMN-----HVGHCYCNLII----------
SP11-54_rapa/1-64                 VE----ACIC----SESF-KGGC-LEP-RN-----------E-YCATQFRKR--LNEQTAFNCTCVT---SH------SRAQCSCRLARKCTTGPWPC---------
SP11-55_rapa/1-57                 TK----NCFI----NGTF-EGAC-ANP-RK-----------E-HCAELVKTR--CNETTAFNCNCGG---SR------TRSHCICQLRRKC----------
SP11-56_rapa/1-53                 KD----GCRD----DFNF-PGIC-DNSGH------------K-KCTDLYVEG--TS-RKHASCVCDGSK---------GVQLCLCNLC----------
SP11-60_rapa/1-62                 AW----KCPEGIVYPSPI-SGRC-INSRS------------T-ECKKHYEVE----GQNVTNCRCDTYSMQ-------NPARITCYCCKVKS----------
SP11-61_rapa/1-52                 ------MMPC----GSIM-VGNC-RNLGA------------K-ECEKLNSPG--KR-KRPSHCKCTNTQR--------GTYSCDCKLC----------
SP11-99_rapa/1-56                 ------LKLC----ARPF-GGRC-GNNGI------------A-LCKMSFGEA--M-NKEAFNCKCEKYNTRNR--------LCKCYFDVHAC----------
SP11-1_oleracea/1-52              ------RCIV----GFRM-PGSC-GNSGD------------K-ECGKLCSPG--KDKKKPSHCKCRNGQE--------DTYACDQCF----------
SP11-3_oleracea/1-52              VMK---NCPI----QFNL-GGQC-GNSGG------------D-ACVEEYNRK--K-KKKKIFCSCGGV--R--------VGQCCCIV----------
SP11-4_oleracea/1-57              LE----ACSR----KERF-EGPC-VDP-RN-----------E-YCAALFKEF--LNENTAFNCTCRT---LV------SRANCRCQLARKC----------
SP11-5_oleracea/1-59              AW----KCPGAVAKADNI-TGTC-VNSVS------------E-DCQRYV-------GQNVNNCLCLNFSKH-------NRGRITCYCCKVKS----------
SP11-6_oleracea/1-51              KK----NCVG----KTRL-PGPC-GDSGA------------S-SCRDLYNQT--E-KTMPVSCRCVP----------TGRCFCSLCK----------
SP11-7_oleracea/1-49              ------QCNI----GYRM-PGNC-AELA-------------Q-TCEKFCSRG--KE-KKPSHWKCTNGPK--------NTYSCDCK----------
SP11-8_oleracea/1-63              MN----KCTD----YINL-LGRC-GGSGD------------G-LCRSSYESN--K-NTKPLNCECKDAKMKFQNDKDVIRGRCRCVLCK----------
SP11-11_oleracea/1-58             MN----QFSC----NRTF-PGAR-CGS-P------------E-DCENYFKHI--LNEKTASNCKCTG---KH------PHLLCTCQLKHKCLP----------
SP11-12_oleracea/1-50             ------PCDD----TFGM-EGQC-GGP--------------K-TCENLYSKG--LD-KRPPRCECTNSGK--------NTYSCVCKLC----------
SP11-13_oleracea/1-51             ------MMPC----GSFM-FGNC-RNIGA------------R-ECEKLNSPG--K--RKPSHCKCTDTQM--------GTYSCDCKLC----------
SP11-13b_oleracea/1-52            ------MMPC----GSFM-FGNC-RNIGA------------R-ECEKLNSPG--KR-KKPSHCKCTDTQM--------GTYSCDCKLC----------
SP11-14_oleracea/1-50             FMK---KCAS----SFPL-SGPC-RQTGV------------K-NCERIYKK-------KPSTCTCEDFFAD-------NNGRCCCV----------
SP11-16_oleracea/1-49             NLMK--ECG-----RFRL-SGLC-GRSGV------------E-DCKIMYKK-------KPSKCTCVNFP-D-------KIGRCCCV----------
SP11-18_oleracea/1-53             MN----RCID----QLPF-RGTC-TSSGG------------E-DCRKLFATE--TN-MHPSRCKCIPDY---------KRRYCRCKLC----------
SP11-20_oleracea/1-49             MIK---KCPD----PINL-RGKC-SESGG------------VVACAKSYNRK------NPSGCSCIDY--D-------EKGRCCL----------
SP11-24_oleracea/1-56             TK----LCPG----NVTS-RGVC-GNSGV------------Q-SCVTAISRK--L-HKDRRLCSCLCKIHE-------GYRFCPCVCKC----------
SP11-25_oleracea/1-64             ---------C----KGTF-RRRC-GSPTNDYCGKLLRDGQNE-KCAKYFREN--LHVKAAFDCQCLY---LH------PHGVCTCQLLRKC----------
SP11-28_oleracea/1-64             VE----ACIC----SESF-EGGC-LEP-RN-----------E-YCAAQYRKR--LNEQTAFNCTCVT---SH------SRAQCSCRLRRKCTTGPWPC---------
SP11-29_oleracea/1-59             LRK---RCPE----HYSL-PGVC-GNSGN------------E-ECKRRYPYPINK-IDLPTTCKCERSKFH-------KRGLCKCSRNC----------
SP11-32_oleracea/1-46             ------RCTR----GFRK-PGKC-TTLEE------------E-KCKTLYRP---------GQCKCSDSKM--------NTHSCDCKSC----------
SP11-33_oleracea/1-59             MK----QCTE----YLNL-GGPC-GNSGK------------A-TCIKLYENI--T-KTTPSYCKCKDLVEKVK-----ITGRCDCRKRKC----------
SP11-39_oleracea/1-57             LN----NCKD----ALRL-PGLC-YSGT-------------L-TCEGLYHKV--C-GRKARSCNCFNDFSNMN-----HVGHCYCNLII----------
SP11-51_oleracea/1-47             ------KEQC----RKNF-PGHC-ETSE---------------RCENTYKRL----NKKVFNCHCQPFGR---------RRLCTCKC----------
SP11-52_oleracea/1-50             KK----DCKD----DIRL-SGLC-HKPG-------------A-SCETVYHQN--C-KTKAKKCNCLNI--D-------GAGRCDCY----------
SP11-57_oleracea/1-49             WMK---GCDG----RIRL-RQLC-GPSGK------------E-RCESINKK-------KPSKCTCVDFD-D-------KIGRCCCL----------
SP11-58_oleracea/1-50             VKE---ECAP----YFGL-NGLC-SQSGE------------N-PCAHQANM-------NRCKCTCGNH--R-------GKGQCFCRVK----------
SP11-61_oleracea/1-52             RN----ECVG----HIRL-SGPC-GDSGE------------N-ACVIRFQRR--E-KKKPRSCGCTNFEG---------IGRCFCCM----------
SP11-62_oleracea/1-51             ------VKRC----DDTF-PGLC-RNNGN------------K-VCEDLFSKH--R-GQKVFNCDCQLFTAKKR--------LCKCKC----------
SP11-64_oleracea/1-53             MK----NCKA----DLRL-PGGC-GNSR-------------I-SCEVLYHDN--C-QRMPSKCKCSNE--S-------DGGRCVCYLII----------
SP11-68_oleracea/1-56             TK----LCRG----SVTS-RGVC-DNSGV------------Q-RCVTAFSKK--I-DKDPRLCSCICRHHD-------GRRFCPCECKC----------
SP11-A14_napus/1-48               MRK---NCPD----GFTL-GGVC-GISGK------------S-ACISSYKNK------KKSDCSCGDV--D-------GIGTCCC----------
SP11-1_sativus/1-54               ------MVPC----DTSF-PGGC-GSGNNGK----------R-VCERSYGV-----ANKPSNCTCVPFIDNNR--------LCKCVVRLC----------
SP11-2_sativus/1-53               VLG---TCDY----TFLL-GGPC-GKSGG------------DKACDKSRTNK-----KNPYDCKCVHT--V-------PKGTCCCGLK----------
SP11-4_sativus/1-48               --------IQ----GFGM-RGKC-GDFGA------------K-KCESLYLAD--KG-KKPSGCKCTDSPN--------NNYVCDCK----------
SP11-6_sativus/1-53               MK----RCTH----QLPF-RGTC-GSSGD------------E-VCKKLYSAE--TK-TNPSRCECIPDY---------KNRFCRCKLC----------
SP11-8_sativus/1-56               ------AKRC----NKTF-NGVC-YKVGAD-----------K-QCEDSYKSA----MKTPYNCTCERFNRQRR--------VCKCNIKEPEC----------
SP11-9_sativus/1-51               AW----NCPEGVTISNPI-LGRC-FNSRS------------R-QCG--------------ENCRCCDYSK--------HNRRITCYCCKVKL----------
SP11-11_sativus/1-44              AP----KCRNGLPEEHAI-KGRC-INSRH------------P-ECNA-------------SNCQCENSNDQ-------KSVRN----------
SP11-15_sativus/1-54              LDVGAWKCPGDVAKSDHI-TGTC-INSKS------------T-DCQQYF-------GQNVTNCVCYTFSKN-------SRGRI----------
SP11-17_sativus/1-70              VK----PCFR----RRTL-TREC-VGYRD------------K-RCDDYFKNK--LNEKTAFNCTCVP---SR------KHALCTCQLRLTPCPYVEDSATSDLI---
SP11-18_sativus/1-50              NMRK--KCH-----FFGL-SGRC-GTSGP------------N-ACAGSYTTK-----KNPSDCNCDDR--R-------KKGKCCCV----------
SP11-19_sativus/1-50              KIK---MCPD----PIKL-RGKC-SESGG------------LAACAKSFKRK------NTAGCSCDDH--D-------EKGSCCCL----------
SP11-21_sativus/1-57              RK----TCVH----RLNT-GGSC-GKSGG------------H-DCEAYYTNK--T-KKQAFYCNCTSPFR---------TRYCDCAVKCKYG----------
SP11-22_sativus/1-56              KK----ECKVLP--FSLV-RKTC-GNSGE------------T-ECKKLYALN--TL-KTASHCRCKIFN---------DSHVCHCVTKC----------
SP11-23_sativus/1-57              VE----ACFR----KEPF-DGRC-IDP-RN-----------E-YCVTLFKEV--LNEKTAFNCTCQT---LH------SRANCRCQLARKC----------
SP11-26_sativus/1-44              AR----KCPGRIDQEYGI-NGRC-IRSKH------------P-DCND-------------SNCYCETSKNQ-------TSWRN----------
SP11-30_sativus/1-53              KN----GCID----ELPF-PGKC-GNSGD------------K-DCIKLYSSE--KN-IHPSKCECEPRY---------KARFCRCKIC----------
SP11-31_sativus/1-75              MK----QCSC----NGTF-LGRC-DDTRNRN----------K-FCADIFSKI--YNEKTAFNCRCSD---AF------PHALCFCQIYVRPGPSCTQKKKVKCIDVV
SP11-1_raphanistrum/1-49          ATK---NCAD----RFSL-GGPC-GKAGA------------V-ACAKVSKKK-----KKPSDCSCEPG--D-------VVGSCCC----------
SP11-6_raphanistrum/1-59          AW----KCPGDVAKSDHI-TGTC-INSKS------------T-DCQQYF-------GQNVTNCVCFTFSKN-------SRGRISCYCCKVKS----------
SP11-4_raphanistrum/1-46          ------RCTR----GFRR-YGAC-DAFEK------------E-KCKKLYPP---------GDCTCSKSKL--------NTHSCDCKSC----------
SP11-4_halleri/1-58               SK----ECNL----RDIF-PGKC-EHDANAK----------L-RCKEDIAKN--FRPSRPFECDCQTFDK---------GRICYCKKCLV----------
SP11-12_halleri/1-54              KK----KCTD----SQVF-GGVC-AFNRN------------Q-GCIADFARD--FK-KKAWDCECINTVH--------NAHLCRCKVC----------
SP11-13_halleri/1-54              KK----KCFE----AKVFPSGKC-GRDGN------------Y-NCKKEYKNS----RVKPTNCKCFSNFD--------ETRLCNCTFC----------
SP11-15_halleri/1-52              RQ----NCIS----KKKF-KGEC-GENGL------------K-VCMEDFKKK----PLKCIKCLDYKPLE---------IHHCTCRFC----------
SP11-20_halleri/1-60              AQ----LCII----NQTF-TGTC-GNNGN------------K-VCIDAVKGR--KRYGTPNQCGCED---AE------SLIFCRCQVNRVHCKR----------
SP11-28_halleri/1-53              NE----GCMR----TESY-TGLCPPRTGN------------I-LCEAEYNNS----KEKPFRCYCRNQI---------GKRVCHCFFC----------
SP11-32_halleri/1-55              KK----LCNL----DQHF-DGTC-GADGN------------K-VCIDEIKSR--NKFNTPNHCGCTN---LG------SSIFCQCKLPC----------
SP11-6_lyrata/1-53                ------NCMW----AGKF-YGPCPLRNAG------------L-SCAAEFSKR--KSKEKPFNCYCANQG---------KWRVCRCLFC----------
SP11-14_lyrata/1-53               ------NCTK----TRTF-RGPCPIRNGG------------S-LCAAQLSNR--KYLVQPFLCRCASHE---------DSRTCYCQVC----------
SP11-16_lyrata/1-62               K-----ACLI----KQIF-PGSC-RPDGY------------I-RCKNDITKN--G-KQRPLECECKDVDG---------DRLCFCYKCLVLTTSDLTT---------
SP11-18_lyrata/1-50               NS----DCIR----TRNL-RGPCPSGDGD------------S-FCRAVFN-------YQVKNCRCLPHW---------DRQVCYCRIC----------
SP11-25_lyrata/1-56               WK----ACRL----RETF-SGTC-GHDGE------------I-RCKNDITRN--GGSPLPFECHCEEFRR---------KRVCHCDKCLL----------
SP11-36_lyrata/1-44               ------ECIV----RDNG-PGIC-GTDGE------------K-VCEEALKAK----QKKPFGCRCKYHLE---------REY----------
SP11-37_lyrata/1-58               WK----ECNL----RDIF-PGKC-EHDANAK----------L-RCKEDIAKN--FRPSRPFECDCQTFDQ---------GRICYCKKCLV----------
SP11-a_lyrata/1-57                KK----VCNL----EGNF-PGRC-VGNGD------------E-QCKRDLTED--G--NNPSKCRCRFRA---------GRRHCRCIYCEVFGM----------
SP11-b_lyrata/1-54                KN----KCMR----SEMFPTGPC-GNNGE------------E-TCKKDFKNI----YRTPIQCKCLDKYD--------FARLCDCRFC----------
```

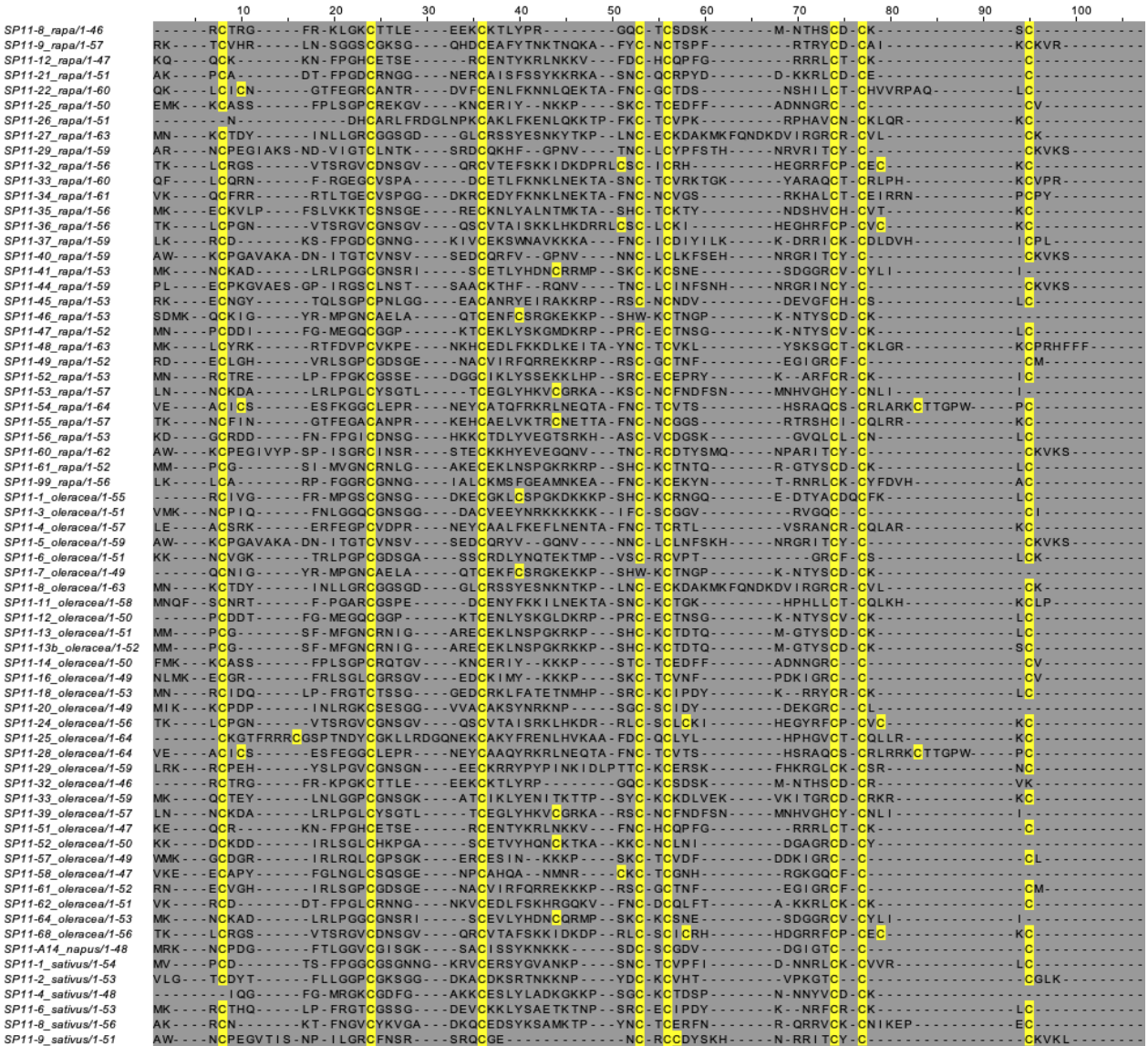



**Figure S3. "seq_MSA" (upper) and "cys-seq MSA" (bottom).** Cysteine residues are highlighted in yellow. This alignment has been visualized using Jalview 2.11.2.6.

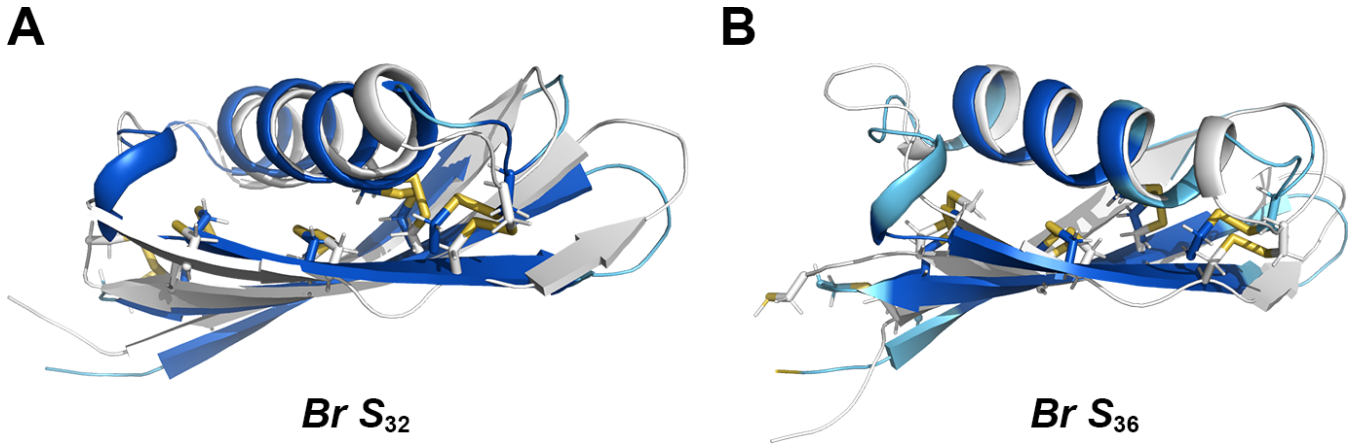



**Figure S4. Superposition of the SP11 models predicted in our previous and current study.** Panels A and B show the predicted models of *Br* $S_{32}$- and $S_{36}$-SP11, respectively. The best models predicted by ColabFold in this study are colored using the pLDDT values. The models from our previous study are shown in white.

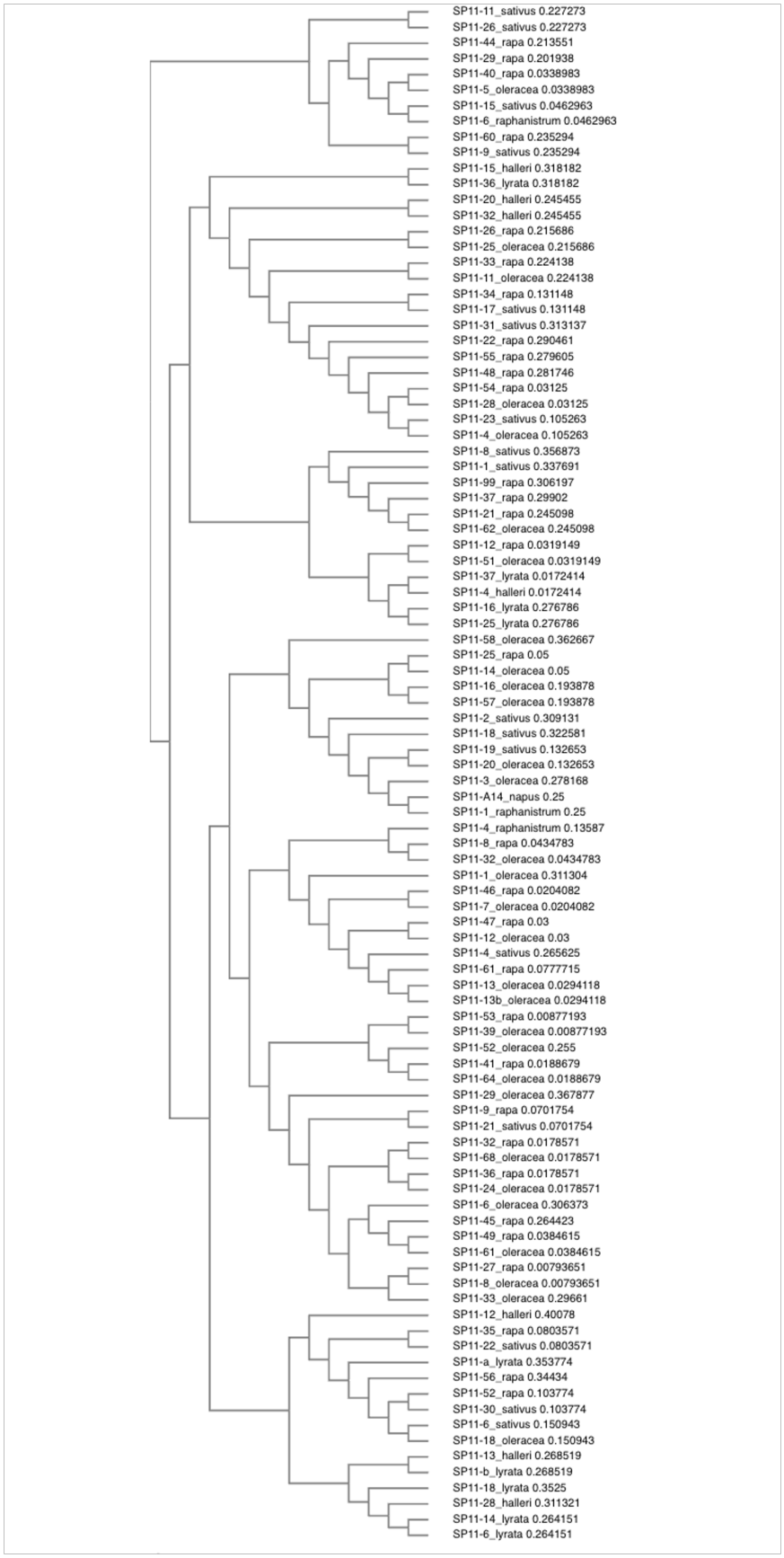



**Figure S5. Phylogenetic tree of the 98 SP11 sequences.** This figure has been created using Clustal Omega 1.2.4.

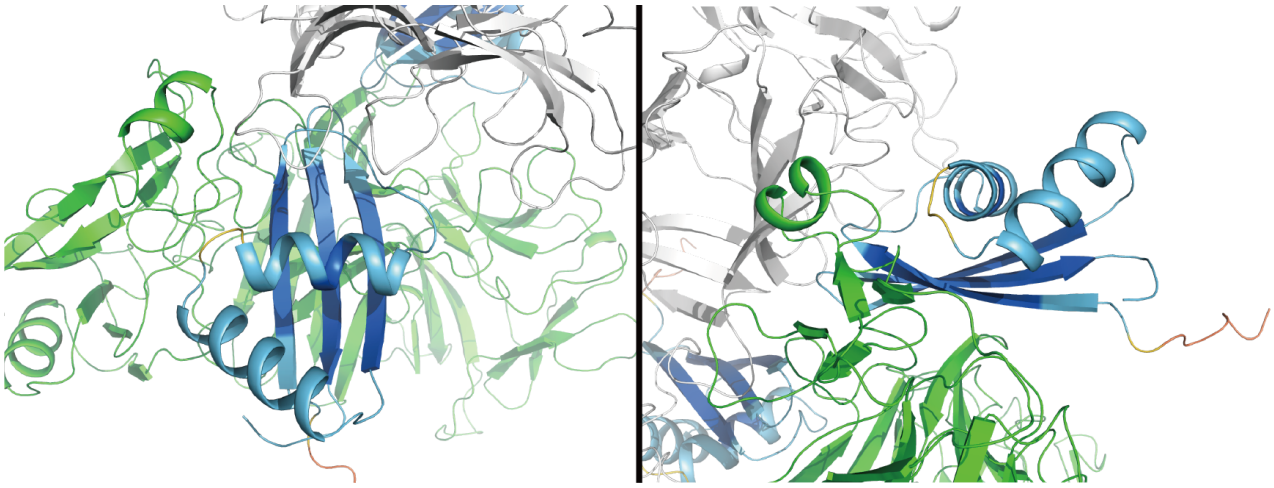


**Figure S6. A predicted model of *Br* $S_{26}$-SRK–SP11.** Close-up view of the interface where the two SRK molecules (white and green) and one SP11 molecule make contact. The second helix of SP11 does not make contact with the SRK molecules.

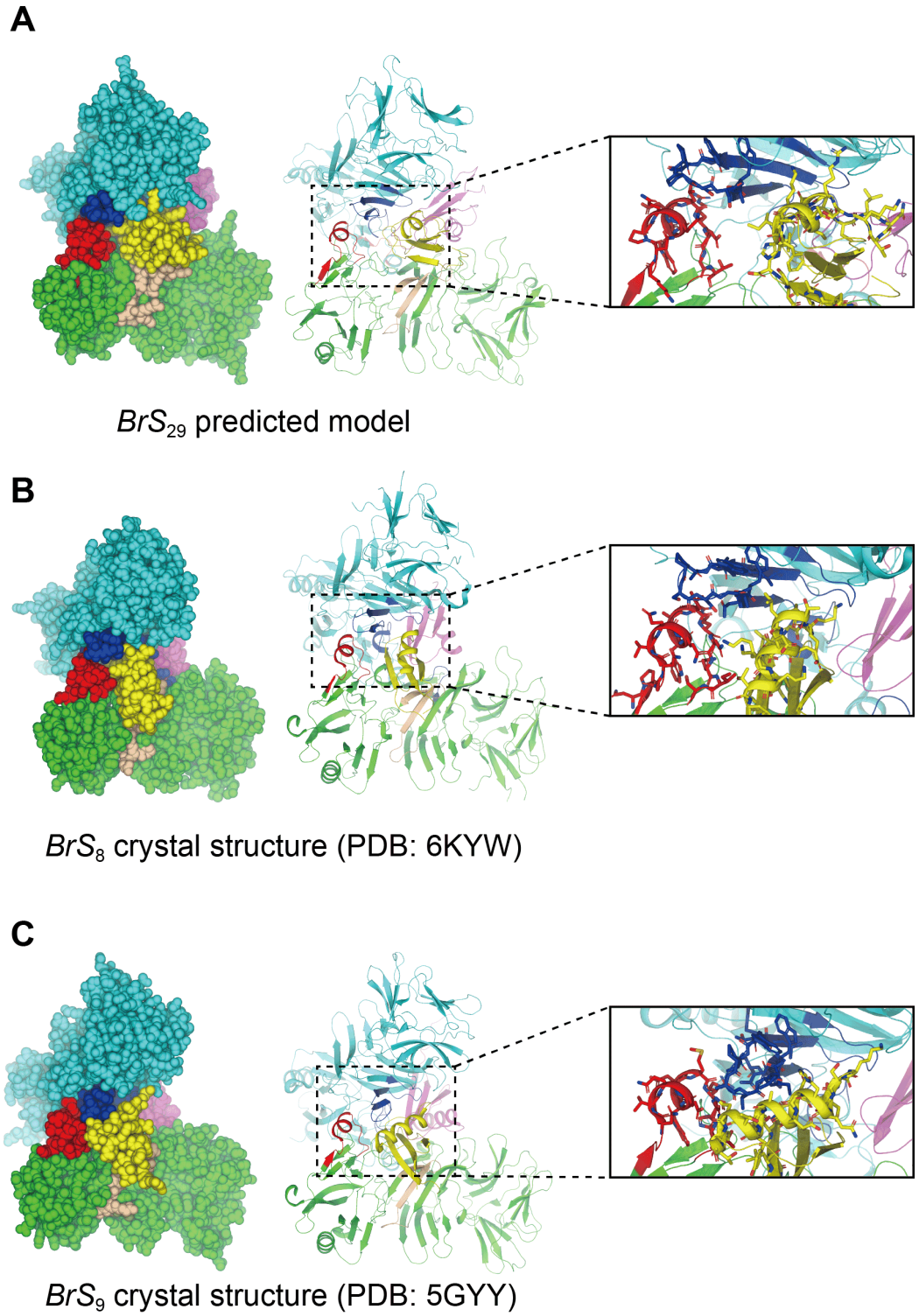
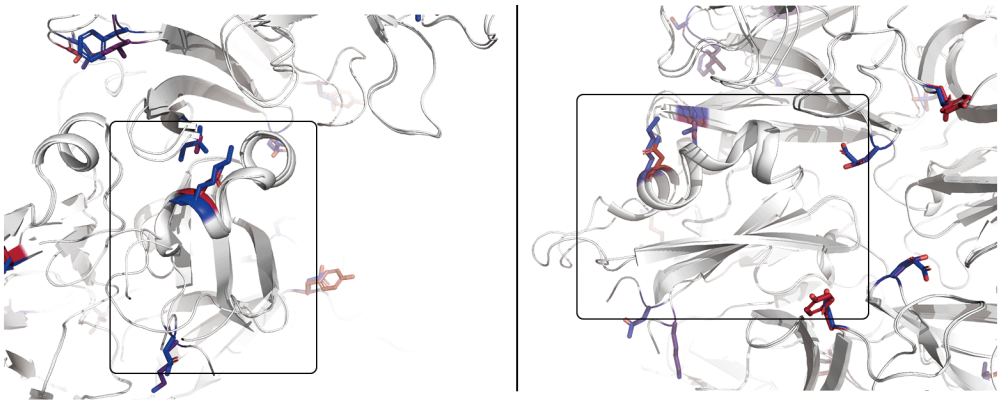
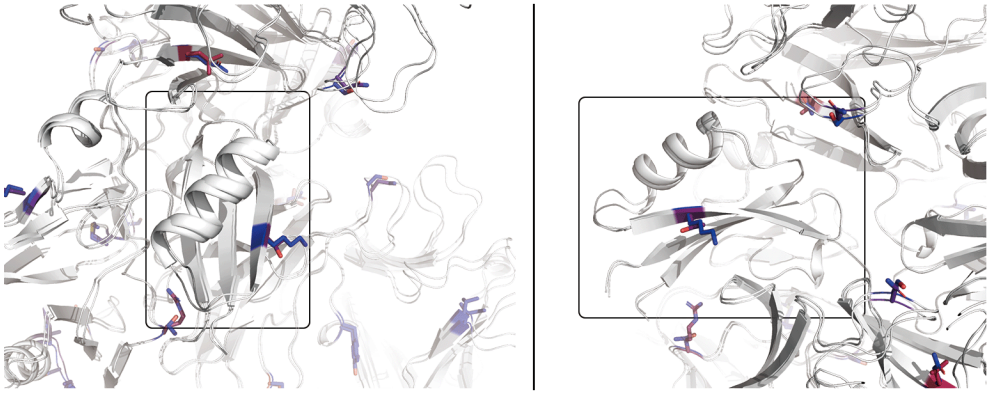



**Figure S7. Structural comparison of the binding interface between HV-III of SRK and SP11.** (A) A predicted model of *Br* $S_{29}$ as a representative of the class-II eSRK–SP11 complexes. (B) The crystal structure of *Br* $S_8$ (PDB: 6KYW). (C) The crystal structure of *Br* $S_9$ (PDB: 5GYY). The two eSRK monomers have been colored in green and cyan and the two SP11s in yellow and purple. In this figure, the position and orientation of SRK chain A (green) are aligned. The HV regions I, II, and III of eSRK are colored in red, deep blue, and light brown, respectively. HV-I and III belong to the same eSRK 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.

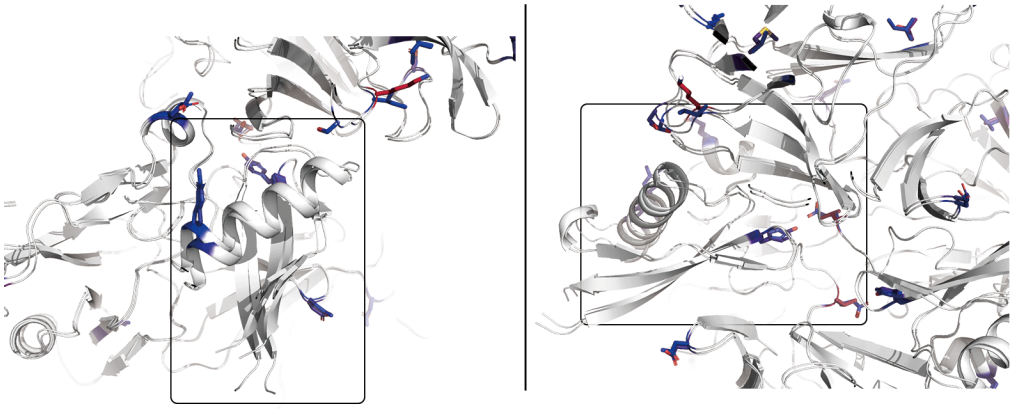
**A** Bo*S*$_{7}$ & Br*S*$_{46}$



**B** Bo*S*$_{32}$ & Br*S*$_{8}$



**C** Bo*S*$_{36}$ & Br*S*$_{24}$



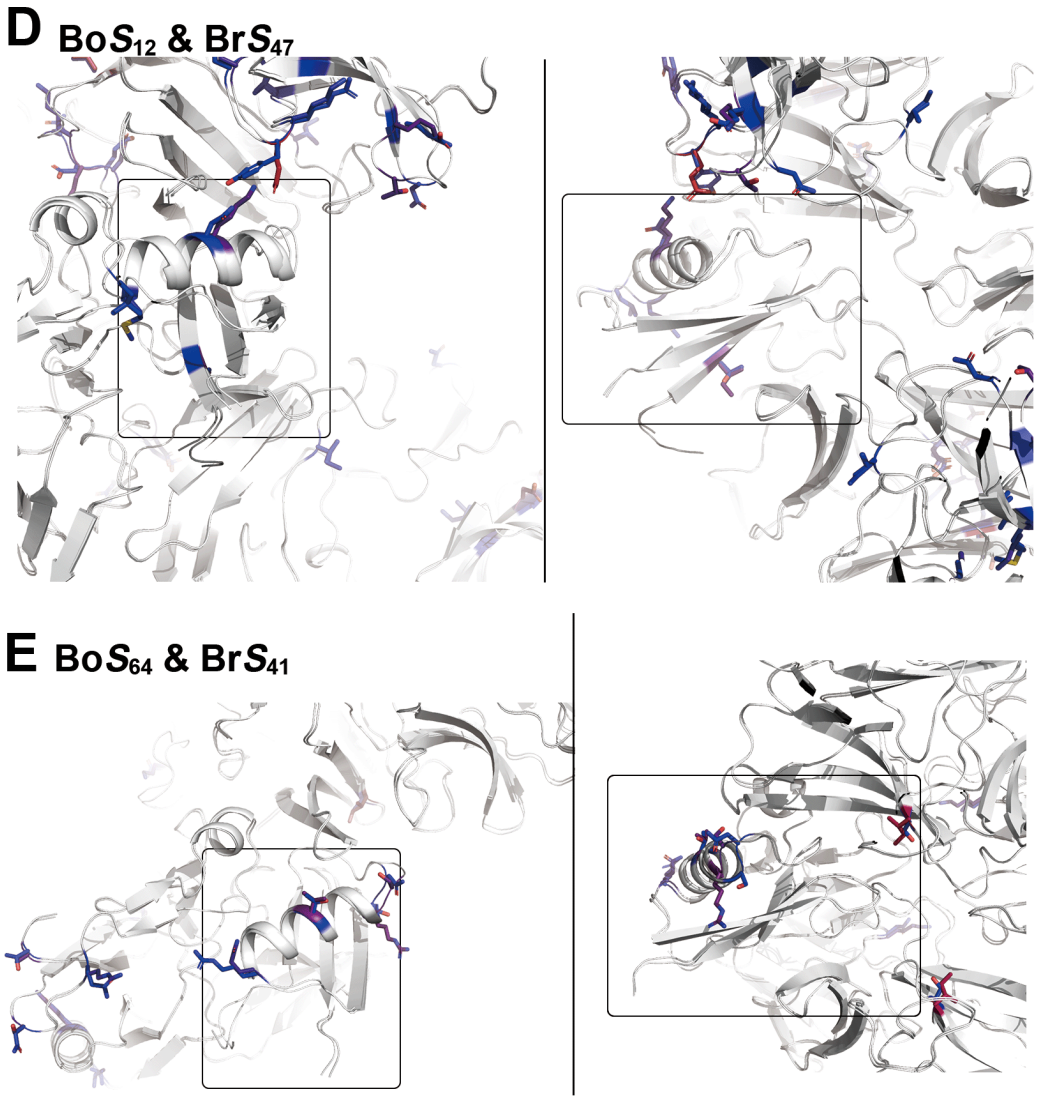
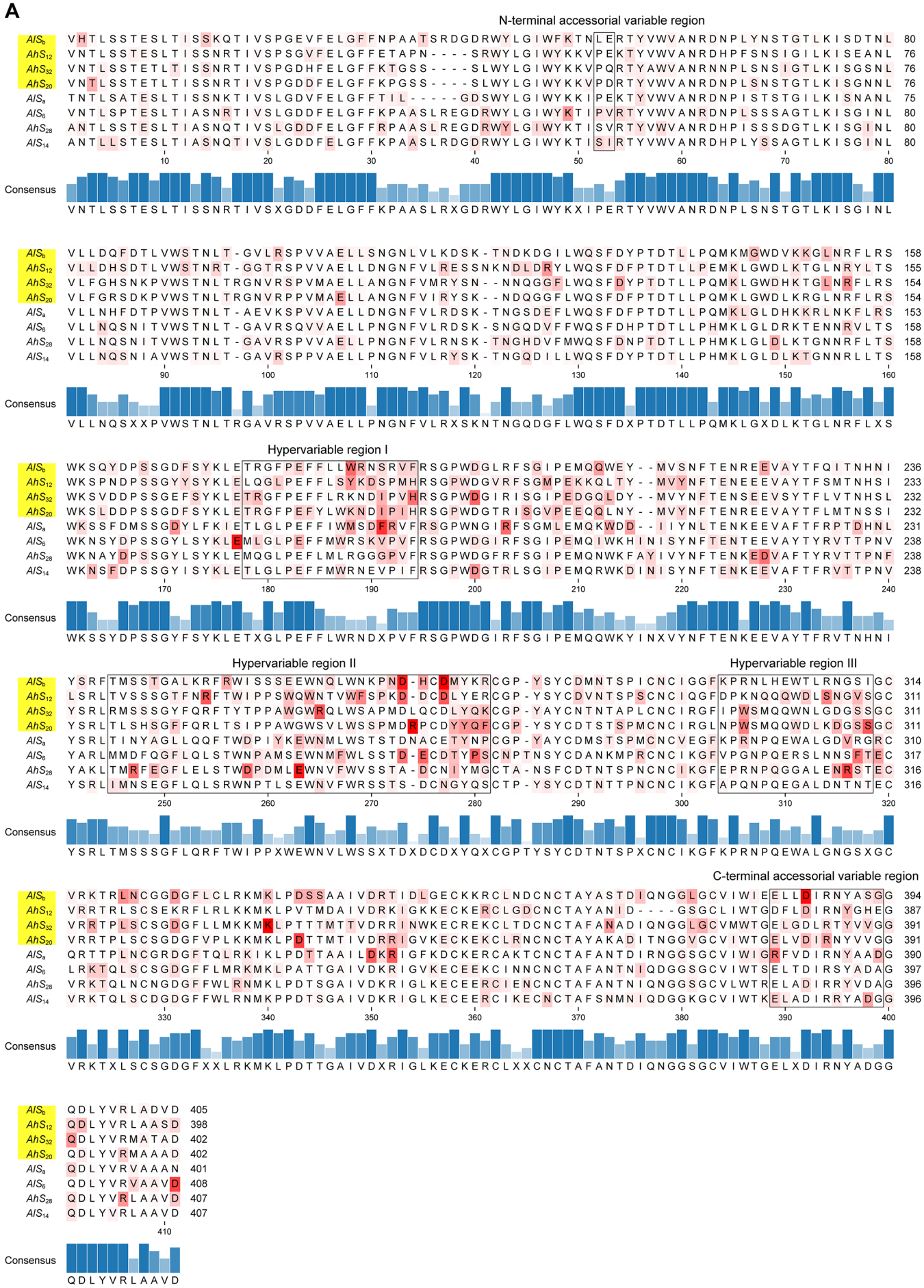



**Figure S8. Superposition of the predicted eSRK–SP11 complex models that have highly similar sequence identity.** (A) *Bo* $S_7$ and *Br* $S_{46}$, (B) *Bo* $S_{32}$ and *Br* $S_8$, (C) *Bo* $S_{36}$ and *Br* $S_{24}$, (D) *Bo* $S_{12}$ and *Br* $S_{47}$, and (E) *Bo* $S_{64}$ and *Br* $S_{41}$. Mutated residues are shown as sticks and in color according to their difference in the BLOSUM90 matrix. The position of SP11 is highlighted inside a black rectangle.

**A**


N-terminal accessorial variable region
AlSb VHTLSSTESLTISSKQTIVSPGEVFELGFFNPAATSRDGDRWYLGIWFKTNLERTYVWVANRDNPLYNSTGTLKISDTNL 80
AhS12 VNTLSSTESLTISSNRTIVSPSGVFELGFFETAPN----SRWYLGIWYKKVPEKTYVWVANRDHPFSNSIGILKISEANL 76
AhS32 VNTLSSTETLTISSNRTIVSPGDHFELGFFKTGSS----SLWYLGIWYKKVPQRTYAWVANRNNPLSNSIGTLKISGSNL 76
AhS20 VNTLSSTETLTISSNRTIVSPGDDFELGFFKPGSS----SLWYLGIWYKKVPDRTYVWVANRDNPLSNSTGTLKISGNNL 76
AlSa TNTLSATESLTISSNKTIVSLGDVFELGFFTIL-----GDSWYLGIWYKKIPEKTYVWVANRDNPISTSTGILKISNANL 75
AlS6 VNTLSPTESLTIASNRTIVSLGDDFELGFFKPAASLREGDRWYLGIWYKTIPVRTYVWVANRDNPLSSSAGTLKISGINL 80
AhS28 ANTLSSTESLTIASNQTIVSLGDDFELGFFRPAASLREGDRWYLGIWYKTISVRTYVWVANRDHPISSSDGTLKISGINL 80
AlS14 ANTLLSTESLTIASNQTIVSLGDDFELGFFKPAASLRDGDRWYLGIWYKTISIRTYVWVANRDHPLYSSAGTLKISGINL 80
10 20 30 40 50 60 70 80
Consensus
VNTLSSTESLTISSNRTIVSXGDDFELGFFKPAASLRXGDRWYLGIWYKXIPERTYVWVANRDNPLSNSTGTLKISGINL
AlSb VLLDQFDTLVWSTNLT-GVLRSPVVAELLSNGNLVLKDSK-TNDKDGILWQSFDYPTDTLLPQMKMGWDVKKGLNRFLRS 158
AhS12 VLLDHSDTLVWSTNRT-GGTRSPVVAELLDNGNFVLRESSNKNDLDRYLWQSFDFPTDTLLPEMKLGWDLKTGLNRYLTS 155
AhS32 VLFGHSNKPVWSTNLTRGNVRSPVMAELLANGNFVMRYSN--NNQGGFLWQSFDYPTDTLLPQMKLGWDHKTGLNRFLRS 154
AhS20 VLFGRSDKPVWSTNLTRGNVRPPVMAELLANGNFVIRYSK--NDQGGFLWQSFDFPTDTLLPQMKLGWDLKRGLNRFLRS 154
AlSa VLLNHFDTPVWSTNLT-AEVKSPVVAELLDNGNFVLRDSK-TNGSDEFLWQSFDFPTDTLLPQMKLGLDHKKRLNKFLRS 153
AlS6 VLLNQSNITVWSTNLT-GAVRSQVVAELLPNGNFVLRDSK-SNGQDVFFWQSFDHPTDTLLPHMKLGLDRKTENNRVLTS 158
AhS28 VLLNQSNITVWSTNLT-GAVRSPVVAELLPNGNFVLRNSK-TNGHDVFMWQSFDNPTDTLLPHMKLGLDLKTGNNRFLTS 158
AlS14 VLLNQSNIAVWSTNLT-GAVRSPPVAELLPNGNFVLRYSK-TNGQDILLWQSFDYPTDTLLPHMKLGLDLKTGNNRLLTS 158
90 100 110 120 130 140 150 160
Consensus
VLLNQSXXPVWSTNLTRGAVRSPVVAELLPNGNFVLRXSKNTNGQDGFLWQSFDXPTDTLLPQMKLGXDLKTGLNRFLXS
Hypervariable region I
AlSb WKSQYDPSSGDFSYKLETRGFPEFFLLWRNSRVFRSGPWDGLRFSGIPEMQQWEY--MVSNFTENREEVAYTFQITNHNI 236
AhS12 WKSPNDPSSGYYSYKLELQGLPEFFLSYKDSPMHRSGPWDGVRFSGMPEKKQLTY--MVYNFTENEEEVAYTFSMTNHSI 233
AhS32 WKSVDDPSSGEFSYKLETRGFPEFFLRKNDIPVHRSGPWDGIRISGIPEDGQLDY--MVYNFTENSEEVSYTFLVTNHSL 232
AhS20 WKSLDDPSSGDFSYKLETRGFPEFYLWKNDIPIHRSGPWDGTRISGVPEEQQLNY--MVYNFTENREEVAYTFLMTNSSI 232
AlSa WKSSFDMSSGDYLFKIETLGLPEFFIWMSDFRVFRSGPWNGIRFSGMLEMQKWDD--IIYNLTENKEEVAFTFRPTDHNL 231
AlS6 WKNSYDPSSGYLSYKLEMLGLPEFFMWRSKVPVFRSGPWDGIRFSGIPEMQIWKHINISYNFTENTEEVAYTYRVTTPNV 238
AhS28 WKNAYDPSSGYLSYKLEMQGLPEFLMLRGGGPVFRSGPWDGFRFSGIPEMQNWKFAYIVYNFTENKEDVAFTYRVTTPNF 238
AlS14 WKNSFDPSSGYISYKLETLGLPEFFMWRNEVPIFRSGPWDGTRLSGIPEMQRWKDINISYNFTENKEEVAFTFRVTTPNV 238
170 180 190 200 210 220 230 240
Consensus
WKSSYDPSSGYFSYKLETXGLPEFFLWRNDXPVFRSGPWDGIRFSGIPEMQQWKYINXVYNFTENKEEVAYTFRVTNHNI
Hypervariable region II
Hypervariable region III
AlSb YSRFTMSSTGALKRFRWISSSEEWNQLWNKPND-HCDMYKRCGP-YSYCDMNTSPICNCIGGFKPRNLHEWTLRNGSIGC 314
AhS12 LSRLTVSSSGTFNRFTWIPPSWQWNTVWFSPKD-DCDLYERCGP-YSYCDVNTSPSCNCIQGFDPKNQQQWDLSNGVSGC 311
AhS32 YSRLRMSSSGYFQRFTYTPPAWGWRQLWSAPMDLQCDLYQKCGP-YAYCNTNTAPLCNCIRGFIPWSMQQWNLGDGSSGC 311
AhS20 YSRLTLSHSGFFQRLTSIPPAWGWSVLWSSPMDRPCDYYQFCGP-YSYCDTSTSPMCNCIRGLNPWSMQQWDLKDGSSGC 311
AlSa YSRLTINYAGLLQQFTWDPIYKEWNMLWSTSTDNACETYNPCGP-YAYCDMSTSPMCNCVEGFKPRNPQEWALGDVRGRC 310
AlS6 YARLMMDFQGFLQLSTWNPAMSEWNMFWLSSTD-ECDTYPSCNPTNSYCDANKMPRCNCIKGFVPGNPQERSLNNSFTEC 317
AhS28 YAKLTMRFEGFLELSTWDPDMLEWNVFWVSSTA-DCNIYMGCTA-NSFCDTNTSPNCNCIKGFEPRNPQGGALENRSTEC 316
AlS14 YSRLIMNSEGFLQLSRWNPTLSEWNVFWRSSTS-DCNGYQSCTP-YSYCDTNTTPNCNCIKGFAPQNPQEGALDNTNTEC 316
250 260 270 280 290 300 310 320
Consensus
YSRLTMSSSGFLQRFTWIPPXWEWNVLWSSXTDXDCDXYQXCGPTYSYCDTNTSPXCNCIKGFKPRNPQEWALGNGSXGC
C-terminal accessorial variable region
AlSb VRKTRLNCGGDGFLCLRKMKLPDSSAAIVDRTIDLGECKKRCLNDCNCTAYASTDIQNGGLGCVIWIEELLDIRNYASGG 394
AhS12 VRRTRLSCSEKRFLRLKKMKLPVTMDAIVDRKIGKKECKERCLGDCNCTAYANID----GSGCLIWTGDFLDIRNYGHEG 387
AhS32 VRRTPLSCSGDGFLLMKKMKLPTTTMTTVDRRINWKECREKCLTDCNCTAFANADIQNGGLGCVMWTGELGDLRTYVVGG 391
AhS20 VRRTPLSCSGDGFVPLKKMKLPDTTMTIVDRRIGVKECKEKCLRNCNCTAYAKADITNGGVGCVIWTGELVDIRNYVVGG 391
AlSa QRTTPLNCGRDGFTQLRKIKLPDTTAAILDKRIGFKDCKERCAKTCNCTAFANTDIRNGGSGCVIWIGRFVDIRNYAADG 390
AlS6 LRKTQLSCSGDGFFLMRKMKLPATTGAIVDKRIGVKECEEKCINNCNCTAFANTNIQDGGSGCVIWTSELTDIRSYADAG 397
AhS28 VRKTQLNCNGDGFFWLRNMKLPDTSGAIVDKRIGLKECEERCIENCNCTAFANTNIQNGGSGCVLWTRELADIRRYVDAG 396
AlS14 VRKTQLSCDGDGFFWLRNMKPPDTSGAIVDKRIGLKECEERCIKECNCTAFSNMNIQDGGKGCVIWTKELADIRRYADGG 396
330 340 350 360 370 380 390 400
Consensus
VRKTXLSCSGDGFXXLRKMKLPDTTGAIVDXRIGLKECKERCLXXCNCTAFANTDIQNGGSGCVIWTGELXDIRNYADGG
AlSb QDLYVRLADVD 405
AhS12 QDLYVRLAASD 398
AhS32 QDLYVRMATAD 402
AhS20 QDLYVRMAAAD 402
AlSa QDLYVRVAAAN 401
AlS6 QDLYVRVAAVD 408
AhS28 QDLYVRLAAVD 407
AlS14 QDLYVRLAAVD 407
410
Consensus
QDLYVRLAAVD

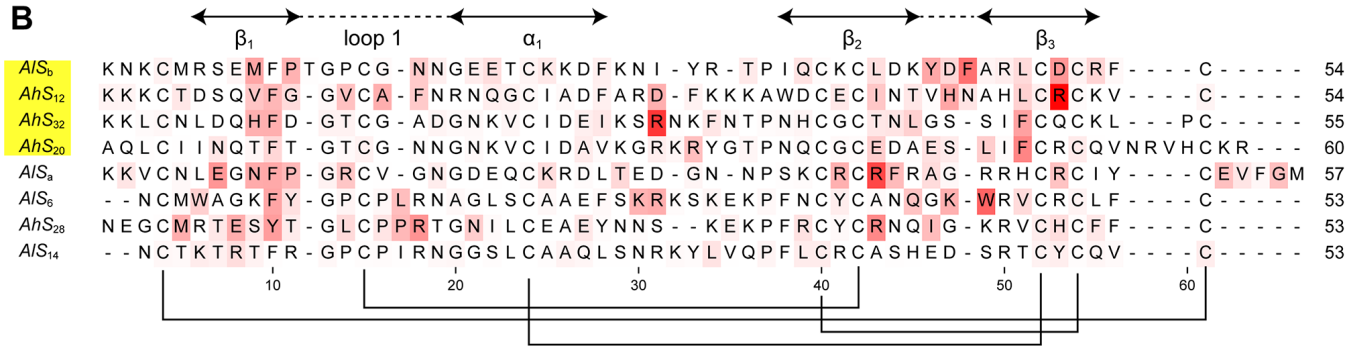



**Figure S9. Per-residue energy decomposition analysis for the predicted *Al*/*Ah* eSRK–SP11 complex models.** (A–B) Sequence alignment of *Al* and *Ah* SRKs (A) and SP11s (B). Haplotypes predicted to have the different binding modes are highlighted in yellow. The variable regions are indicated by black borders.

```
#<length of SRK>,<length of SP11> <Number of SRK monomers>,<Number of SP11 monomers>
>101 102 /* concatenated SRK and SP11 sequence to be predicted */
AAAAAAAAAAAAAAAAAAaaaaaaaaaa
>SRK_B SP11_B
BBBBBBBBBBBBBBBBBBbbbb-bbbbb
>SRK_C SP11_C
CCCC--CCCCCCCCCCCCccc--ccccc
...
/* concatenated SRK and SP11 sequences derived from the same S haplotype */
...
>101 /* SRK sequence to be predicted */
AAAAAAAAAAAAAAAAAA----------
>UPI000BBB67E1 559  1.00 /* SRK_homologs */
LSTLSSTESLTISSNRTL----------
...
/* alignment of SRK homolog sequences */
...
>102 /* SP11 sequence to be predicted */
------------------aaaaaaaaaa
>Q9ST12   91   1.00 /* SP11_homologs */
------------------RKTCVHRLNS
...
/* alignment of SP11 homolog sequences */
...
```

**Figure S10. Sample a3m file for structure prediction of cognate SRK and SP11 complex.** The amino acid length of SRK and SP11, and the number of SRK and SP11 monomers are specified in the first line starting with # character (e.g. 398,57 2,2). Concatenated SRK and SP11 sequence to be predicted are placed in the first row of the first block, followed by the remaining 97 concatenated SRK–SP11 sequence pairs. SRK and SP11 homolog sequence alignments with gaps are placed in the second and third blocks, respectively.