Can AI modelling of protein structures distinguish between sensor and helper NLR immune receptors?

AmirAli Toghani[1], Ryohei Terauchi[2,3], Sophien Kamoun[1*], Yu Sugihara[2*]

[1] The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, United Kingdom
[2] Iwate Biotechnology Research Center, Kitakami, Iwate, Japan
[3] Crop Evolution Laboratory, Kyoto University, Muko, Kyoto, Japan

*Corresponding authors: sophien.kamoun@tsl.ac.uk (S.K) and sugihara.yu.85s@kyoto-u.jp (Y.S.)

ORCID IDs
AmirAli Toghani: 0000-0002-0644-699X
Ryohei Terauchi: 0000-0002-0095-4651
Sophien Kamoun: 0000-0002-0290-0315
Yu Sugihara: 0000-0001-6042-1091

ABSTRACT

NLR immune receptors can be functionally organized in genetically linked sensor-helper pairs. However, methods to categorize paired NLRs remain limited, primarily relying on the presence of non-canonical domains in some sensor NLRs. Here, we propose that the AI system AlphaFold 3 can classify paired NLR proteins into sensor or helper categories based on predicted structural characteristics. Helper NLRs showed higher AlphaFold 3 confidence scores than sensors when modelled in oligomeric configurations. Furthermore, funnel-shaped structures—essential for activating immune responses—were reliably predicted in helpers but not in sensors. Applying this method to uncharacterized NLR pairs from rice, we found that AlphaFold 3 can differentiate between putative sensors and helpers even when both proteins lack non-canonical domain annotations. These findings suggest that AlphaFold 3 offers a new approach to categorize NLRs and enhances our understanding of the functional configurations in plant immune systems, even in the absence of non-canonical domain annotations.

INTRODUCTION

NLR (nucleotide binding and leucine-rich repeat) proteins are intracellular immune receptors that occur across all kingdoms of life but are particularly highly diversified in plants (Barragan and Weigel, 2021). These proteins function as singletons, pairs, or networks (Adachi et al., 2019b; Contreras et al., 2023a). Singleton NLRs can detect pathogens and execute hypersensitive cell death and immune responses, while paired NLRs have subfunctionalized into sensor (pathogen detection) and helper (immune execution) NLRs that carry distinct biochemical activities. In monocots, paired NLRs typically emerge from distinct phylogenetic clades to function together and do not have a common phylogenetic origin, unlike NLR networks (Contreras et al., 2023a; Kourelis et al., 2021). Sensor and helper NLR pairs are often genetically clustered, and some sensors have non-canonical integrated domains (IDs) that function in pathogen sensing and are absent in helper NLRs (Białas et al., 2018; Marchal et al., 2022). However, the presence of IDs is currently the only effective way to distinguish sensor NLRs from helper NLRs *in silico*, and no other method has been established to distinguish sensors from helpers in paired NLRs.

In plants, NLRs that carry a coiled-coil (CC) domain at their N-termini are the most widespread class. Following pathogen recognition, CC-NLR proteins oligomerize into pentameric or hexameric pore-like complexes (Förderer et al., 2022; Liu et al., 2024; Madhuprakash et al., 2024; Wang et al., 2019; Zhao et al., 2022). These complexes, known as resistosomes, are a defining feature of NLRs that execute the immune response such as the helpers; they translocate to cellular membranes and trigger immune responses such as calcium influx and hypersensitive cell death (Contreras et al., 2022; Duggan et al., 2021; Ibrahim et al., 2024). The prevailing model is that the funnel-shaped structure of CC-NLR resistosomes inserts into membranes and is required for cell death and immune response (Adachi et al., 2019b; Förderer and Kourelis, 2023; Wang et al., 2019). However, this funnel-shaped structure is formed by the N-terminal α1 helix of NLR, which is a structurally dynamic region that is difficult to resolve using cryo-electron microscopy (Förderer et al., 2022; Liu et al., 2024; Madhuprakash et al., 2024; Zhao et al., 2022). About 20% of plant CC-NLRs have a conserved sequence motif, called MADA, at the N-terminal α1 helix, while a previous study indicates that MADA motifs have degenerated in some sensor NLRs of solanaceous plants (Adachi et al., 2019a).

AlphaFold revolutionized protein structural biology by accurately modelling protein structures from primary sequences (Jumper et al., 2021). Recently, AlphaFold 3 (AF3) was released, enabling the modelling of protein structures with oleic acids serving as proxy for cellular membranes (Abramson et al., 2024). This function is particularly useful for proteins that dynamically interact with cellular membranes, such as CC-NLRs, and has enabled high-confidence modelling of CC-NLR resistosomes (Madhuprakash et al., 2024; Ibrahim et al., 2024).

Here, we leveraged AF3 to explore the structural diversity of sensor and helper oligomers of a curated set of CC-NLRs consisting of i) experimentally validated NLR pairs in rice (Pikm, Pii, and Pia), ii) their orthologs (PIK5/6-NP, Pi5-3/1, and Pias), and iii) two previously cloned NLR pairs in barley (RPG5/HvRGA1 and RGH2/3) as summarized in **Table S1**. Our analyses revealed that helpers consistently exhibit higher AF3 confidence scores than sensors. Moreover, the funnel-shaped structures, crucial for the immune response, were consistently observed in helpers but not in sensors. We extended our investigations to uncharacterized NLR pairs of rice, and showed that AF3 can distinguish between NLR pairs even when both proteins do not carry any annotations for unconventional integrated domains. Based on these findings, we propose that AF3 can be used to classify paired NLRs into functional categories, providing a novel approach to understanding functional configurations of plant immune receptors.

RESULTS

Helper NLRs produce higher AlphaFold 3 pTM scores than their paired sensors

To benchmark the AF3 capacity to distinguish between sensor and helper NLRs, we curated eight NLR pairs in which the sensor can be readily assigned because it carries an ID (**Fig. S1**; **Table S1**). As in previous studies (Madhuprakash et al., 2024; Ibrahim et al., 2024), we used the oligomerizing domains of the NLR proteins, from the N-terminus to the end of the NB-ARC domain, and performed AF3 predictions of 5x and 6x stoichiometries with 50 oleic acids and using three different seed values. We then compared the sensor and helper predicted template modelling (pTM) scores (**Fig. 1a and b**; **Fig. S2–4**; **Table S2**). Helper NLRs consistently exhibited higher pTM scores than sensor NLRs in both pentameric and hexameric configurations, indicating that AF3 pTM scores can be used to distinguish between sensor and helper NLRs.

Helper AF3 structures form funnel-shaped structures unlike sensors

In addition to the AF3 confidence scores, we also examined the AF3 structures for distinct structural patterns between sensor and helper NLRs (**Fig. 1c**; **Fig. S5 and S6**). Helper NLRs consistently formed funnel-shaped structures in contrast to sensor NLRs. This observation aligns with previous studies demonstrating that sensor NLRs cannot execute the immune response on their own and may have lost the capacity to oligomerise into resistosome-like structures (Adachi et al., 2019b; Contreras et al., 2022). Sensors formed resistosome-like structures but with low confidence scores (**Fig. 1c**; **Fig. S5 and S6**). These findings further indicate that AF3 can distinguish between sensor and helper NLRs.

We then analyzed whether the presence of the MADA motif could classify sensors or helpers in the curated NLR pairs using the MADA HMM model in Adachi et al., 2019a (**Table S3**). The helpers of Pikm and PIK5/6-NP contained the MADA motif with HMM scores exceeding the cut-off value of 10, while the other sensors and helpers did not. Based on these results, only Pikm and PIK5/6-NP could be classified according to the presence of the MADA motif, whereas the remaining NLR pairs could not be classified. These findings indicate that AF3 modelling-based classification is more reliable than the sequence-based classification and highlight the utility of AF3 for NLR classification.

AF3 can classify paired NLRs even when the putative sensors lack an integrated domain annotation

Based on Stein et al., 2018, we extracted ten rice CC-NLR pairs that i) are genetically linked in head-to-head orientations; ii) belong to distinct phylogenetic clades; and iii) carry a full N-terminal CC domain (**Fig. 2a**; **Fig. S1; Table S4–S6**). In five of the ten pairs, one of the NLRs carries an ID annotation and is presumed to be the sensor. The putative helpers (without ID annotation) had higher AF3 pTM scores than sensors (with ID annotation) for these pairs (**Fig. 2b**; **Fig. S7; Table S7**). Furthermore, putative helpers formed funnel-shaped structures, whereas putative sensors did not (**Fig. 2c**; **Fig. S8 and S9**). These results further confirm the application of AF3 for functional classification as noted with the eight previously characterized pairs.

Next, we examined the five remaining pairs whose NLRs do not carry any ID annotations. In each pair, we averaged the pTM scores of pentameric and hexameric configurations and assigned them as putative helpers or sensors based on higher or lower pTM scores, respectively (**Fig. 2b**; **Fig. S1 and S7; Table S7**). Pair-13 is identical to the rice pair PIK5/6-NP, in which PIK5-NP is known as a sensor with an integrated domain that is not annotated by InterProScan (Białas et al., 2021; Kourelis et al., 2021). Nonetheless, AF3 successfully classified them into helper or sensor based on the pTM scores even without the ID annotation. Among the other four pairs, Pair-07 exhibited a consistent difference in the pTM scores between the putative sensor and helper NLRs (**Fig. S7; Table S7**).

DISCUSSION

AlphaFold 3 can potentially distinguish between sensor and helper NLRs based on their predicted structures and confidence scores. In AF3 predictions, helpers consistently exhibited higher AF3 confidence scores than sensors (**Fig. 1**; **Fig. S2–4**). Moreover, the funnel-shaped structures were consistently observed in helpers but not in sensors (**Fig. 1c and 2c**; **Fig. S5, S6, S8, and S9**). These findings indicate that AF3 might be able to distinguish the functional

configurations of plant immune receptors. Our results highlight the utility of AF3 as a powerful tool for distinguishing structural and functional differences in paired NLRs.

Why did sensor NLRs exhibit lower AF3 confidence scores than helper NLRs? Previous studies proposed a model suggesting that paired NLRs evolved from singleton NLRs (Adachi et al., 2019b). In this model, singleton NLRs, which can detect pathogens (sensor) and execute hypersensitive cell death (helper), subfunctionalized into sensor or helper NLRs (Adachi et al., 2019b; Contreras et al., 2023a). This subfunctionalization results in the loss of cell death activity in sensor NLRs and the loss of pathogen perception activity in helper NLRs. Since the formation of resistosomes with funnel-shaped structures is essential for induction of cell death and other immune responses, the presence of stable resistosome and funnel-shaped structures in helper NLRs, but not in sensor NLRs, supports the previously proposed model. Consequently, sensor NLRs exhibit lower AF3 confidence scores than helper NLRs, reflecting their evolutionary divergence and functional specialization.

Recently, several researchers have applied AlphaFold to plant-pathogen interactions (Contreras et al., 2023b; Cruz et al., 2024; Ibrahim et al., 2024, 2023; Madhuprakash et al., 2024; Selvaraj et al., 2024; Seong et al., 2024; Sugihara et al., 2023; Tamborski et al., 2023). These advancements provide a novel approach to understanding functional configurations of immune receptors beyond domain architectures, sequence motifs, and phylogenetic relationships.

METHODS

NLR annotation

NLRs were annotated using NLRtracker v1.0.3 (Kourelis et al., 2021) and InterProScan v5.67-99.0 (Jones et al., 2014). Note that InterProScan did not annotate an integrated domain of PIK5-NP as previously reported (Kourelis et al., 2021). Therefore, we manually replaced the domain architecture of PIK5-NP from "CNL" to "CONL" as it contains the HMA domain between the NB-ARC and LRR (Białas et al., 2021). The domain architectures were visualized using refplantnlR (https://github.com/JKourelis/refplantnlR). The MADA motifs were analyzed using the HMM model in Adachi et al., 2019a and HMMER v3.4 (http://hmmer.org) with the option "--max". The NLRtracker and HMMER outputs are archived on Zenodo (https://doi.org/10.5281/zenodo.13826775).

Curation of NLR sequences

The sequences of curated NLR pairs were derived from either RefPlantNLR (Kourelis et al., 2021) or NCBI (**Table S1**). Regarding the NLR pairs described by Stein et al., 2018, the protein sequences of *Oryza sativa* cv. Nipponbare (Oryza_sativa_vg_japonica.protein.fasta) were downloaded from the URL (https://doi.org/10.7946/P2FC9Z). Based on Supplementary Data 6 in Stein et al., 2018, the NLR pairs that i) are genetically linked in head-to-head orientations; ii) belong to distinct phylogenetic clades were extracted (**Table S4**). Using DIAMOND BLASTP v2.1.9 (Buchfink et al., 2021), the corresponding NLR sequences were identified from the NCBI RefSeq annotation of rice cultivar Nipponbare genome (GCF_034140825.1). The best-hit sequences were summarized in **Table S5** and confirmed to be genetically linked in a head-to-head orientation. Based on NLRtracker outputs, the CC-NLR pairs, which carry a full N-terminal CC domain, were retained (**Table S6**). Piar-05, Piar-13, and the sensor of Pair-14 correspond to Pia, PIK5/6-NP, and Pi-ta, respectively (**Table S4 and S6**).

AF3 prediction

Based on NLRtracker outputs, the amino acid sequences from the N-terminus to the end of the NB-ARC domain were extracted. Using the AlphaFold 3 webserver (https://golgi.sandbox.google.com), the extracted sequences were modelled in both pentameric and hexameric configurations with 50 oleic acids using three different seed values. The input sequences and resulting models are archived on Zenodo (https://doi.org/10.5281/zenodo.13826775).

Phylogenetic analysis

We built the tree of monocot paired NLRs with RefPlantNLR (Kourelis et al., 2021) and a set of 4,936 NLR proteins from the NLRtracker output of 13 RefSeq proteomes, including two dicot species (*Arabidopsis thaliana* and *Solanum lycopersicum*) and 11 monocot species (*Zea mays*, *Triticum aestivum*, *Setaria viridis*, *Phragmites australis*, *Phoenix dactylifera*, *Oryza sativa*, *Musa acuminata*, *Lolium perenne*, *Hordeum vulgare* subsp. *vulgare*, *Brachypodium distachyon*, and *Asparagus officinalis*) (**Data S1, S2, S3, and S4**) (Toghani and Kamoun, 2024). More details on the phylogenetics analysis are available on GitHub (https://github.com/amiralito/Paired_NLR_AF3).

FUNDING

The authors received funding from the sources listed below. The Gatsby Charitable Foundation (A.T. and S.K.), Biotechnology and Biological Sciences Research Council (BBSRC) BB/P012574 (Plant Health ISP) (AT and SK), BBSRC BBS/E/J/000PR9795 (Plant Health ISP - Recognition)

(A.T. and S.K.), BBSRC BBS/E/J/000PR9796 (Plant Health ISP - Response) (A.T. and S.K.), BBSRC BBS/E/J/000PR9797 (Plant Health ISP – Susceptibility) (A.T. and S.K.), BBSRC BBS/E/J/000PR9798 (Plant Health ISP – Evolution) (A.T. and S.K.), European Research Council (ERC) 743165 (S.K.), Engineering and Physical Sciences Research Council EP/Y032187/1 (S.K.), and Japan Society for the Promotion of Science (JSPS) 23K20042, 24H00010 (R.T.).

ACKNOWLEDGEMENTS

We thank Tolga Bozkurt (Imperial College London), Andres Posbeyikian (The Sainsbury Laboratory), Raoul Frijters (Rijk Zwaan Breeding B.V.), and Kentato Fujita (Osaka University) for valuable comments and scientific discussions.

CONFLICT OF INTERESTS

S.K. receives funding from industry on NLR biology and has co-founded a start-up company (Resurrect Bio Ltd.) related to NLR biology. S.K. has filed patents on NLR biology.

DATA AVAILABILITY

The scripts and dataset used for the phylogenetic analysis are available on GitHub (https://github.com/amiralito/Paired_NLR_AF3). The other datasets are archived on Zenodo (https://doi.org/10.5281/zenodo.13826775).

REFERENCES

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung C-C, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Žídek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**:493–500. doi:10.1038/s41586-024-07487-w

Adachi H, Contreras MP, Harant A, Wu C, Derevnina L, Sakai T, Duggan C, Moratto E, Bozkurt TO, Maqbool A, Win J, Kamoun S. 2019a. An N-terminal motif in NLR immune receptors is functionally conserved across distantly related plant species. *eLife* **8**:e49956. doi:10.7554/eLife.49956

Adachi H, Derevnina L, Kamoun S. 2019b. NLR singletons, pairs, and networks: evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants. *Curr Opin Plant Biol* **50**:121–131. doi:10.1016/j.pbi.2019.04.007

Barragan AC, Weigel D. 2021. Plant NLR diversity: the known unknowns of pan-NLRomes. *Plant Cell* **33**:814–831. doi:10.1093/plcell/koaa002

Białas A, Langner T, Harant A, Contreras MP, Stevenson CE, Lawson DM, Sklenar J, Kellner R, Moscou MJ, Terauchi R, Banfield MJ, Kamoun S. 2021. Two NLR immune receptors

acquired high-affinity binding to a fungal effector through convergent evolution of their integrated domain. *eLife* **10**:e66961. doi:10.7554/eLife.66961

Białas A, Zess EK, De la Concepcion JC, Franceschetti M, Pennington HG, Yoshida K, Upson JL, Chanclud E, Wu C-H, Langner T, Maqbool A, Varden FA, Derevnina L, Belhaj K, Fujisaki K, Saitoh H, Terauchi R, Banfield MJ, Kamoun S. 2018. Lessons in Effector and NLR Biology of Plant-Microbe Systems. *Mol Plant-Microbe Interactions®* **31**:34–45. doi:10.1094/MPMI-08-17-0196-FI

Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**:366–368. doi:10.1038/s41592-021-01101-x

Contreras MP, Lüdke D, Pai H, Toghani A, Kamoun S. 2023a. NLR receptors in plant immunity: making sense of the alphabet soup. *EMBO Rep* **24**:e57495. doi:10.15252/embr.202357495

Contreras MP, Pai H, Selvaraj M, Toghani A, Lawson DM, Tumtas Y, Duggan C, Yuen ELH, Stevenson CEM, Harant A, Maqbool A, Wu C-H, Bozkurt TO, Kamoun S, Derevnina L. 2023b. Resurrection of plant disease resistance proteins via helper NLR bioengineering. *Sci Adv* **9**:eadg3861. doi:10.1126/sciadv.adg3861

Contreras MP, Pai H, Tumtas Y, Duggan C, Yuen ELH, Cruces AV, Kourelis J, Ahn H-K, Lee K-T, Wu C-H, Bozkurt TO, Derevnina L, Kamoun S. 2022. Sensor NLR immune proteins activate oligomerization of their NRC helpers in response to plant pathogens. *EMBO J* **n/a**:e111519. doi:10.15252/embj.2022111519

Cruz DGDL, Zdrzałek R, Banfield MJ, Talbot NJ, Moscou MJ. 2024. Molecular mimicry of a pathogen virulence target by a plant immune receptor. doi:10.1101/2024.07.26.605320

Duggan C, Moratto E, Savage Z, Hamilton E, Adachi H, Wu C-H, Leary AY, Tumtas Y, Rothery SM, Maqbool A, Nohut S, Martin TR, Kamoun S, Bozkurt TO. 2021. Dynamic localization of a helper NLR at the plant–pathogen interface underpins pathogen recognition. *Proc Natl Acad Sci* **118**:e2104997118. doi:10.1073/pnas.2104997118

Förderer A, Kourelis J. 2023. NLR immune receptors: structure and function in plant disease resistance. *Biochem Soc Trans* **51**:1473–1483. doi:10.1042/BST20221087

Förderer A, Li E, Lawson AW, Deng Y, Sun Y, Logemann E, Zhang X, Wen J, Han Z, Chang J, Chen Y, Schulze-Lefert P, Chai J. 2022. A wheat resistosome defines common principles of immune receptor channels. *Nature* **610**:532–539. doi:10.1038/s41586-022-05231-w

Ibrahim T, Khandare V, Mirkin FG, Tumtas Y, Bubeck D, Bozkurt TO. 2023. AlphaFold2-multimer guided high-accuracy prediction of typical and atypical ATG8-binding motifs. *PLOS Biol* **21**:e3001962. doi:10.1371/journal.pbio.3001962

Ibrahim T, Yuen ELH, Wang H-Y, King FJ, Toghani A, Kourelis J, Vuolo C, Adamkova V, Castel B, Jones JD, Wu C-H, Kamoun S, Bozkurt TO. 2024. A helper NLR targets organellar membranes to trigger immunity. doi:10.1101/2024.09.19.613839

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240. doi:10.1093/bioinformatics/btu031

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, 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 AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. doi:10.1038/s41586-021-03819-2

Kourelis J, Sakai T, Adachi H, Kamoun S. 2021. RefPlantNLR is a comprehensive collection of experimentally validated plant disease resistance proteins from the NLR family. *PLOS Biol* **19**:e3001124. doi:10.1371/journal.pbio.3001124

Liu F, Yang Z, Wang C, You Z, Martin R, Qiao W, Huang J, Jacob P, Dangl JL, Carette JE, Luan S, Nogales E, Staskawicz BJ. 2024. Activation of the helper NRC4 immune receptor forms a hexameric resistosome. *Cell* **187**:4877-4889.e15. doi:10.1016/j.cell.2024.07.013

Madhuprakash J, Toghani A, Contreras MP, Posbeyikian A, Richardson J, Kourelis J, Bozkurt TO, Webster M, Kamoun S. 2024. A disease resistance protein triggers oligomerization of its NLR helper into a hexameric resistosome to mediate innate immunity.

Marchal C, Michalopoulou VA, Zou Z, Cevik V, Sarris PF. 2022. Show me your ID: NLR immune receptors with integrated domains in plants. *Essays Biochem* EBC20210084. doi:10.1042/EBC20210084

Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. 2023. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* **32**:e4792. doi:10.1002/pro.4792

Selvaraj M, Toghani A, Pai H, Sugihara Y, Kourelis J, Yuen ELH, Ibrahim T, Zhao H, Xie R, Maqbool A, Concepcion JCD la, Banfield MJ, Derevnina L, Petre B, Lawson DM, Bozkurt TO, Wu C-H, Kamoun S, Contreras MP. 2024. Activation of plant immunity through conversion of a helper NLR homodimer into a resistosome. *PLOS Biol* **22**:e3002868. doi:10.1371/journal.pbio.3002868

Seong K, Wei W, Vega B, Dee A, Ramirez-Bernardino G, Kumar R, Parra L, Krasileva K. 2024. Engineering the plant intracellular immune receptor Sr50 to restore recognition of the AvrSr50 escape mutant. doi:10.1101/2024.08.07.607039

Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoechea JL, Wei S, Wang J, Liao Y, Wang M, Jacquemin J, Becker C, Kudrna D, Zhang J, Londono CEM, Song X, Lee S, Sanchez P, Zuccolo A, Ammiraju JSS, Talag J, Danowitz A, Rivera LF, Gschwend AR, Noutsos C, Wu C, Kao S, Zeng J, Wei F, Zhao Q, Feng Q, El Baidouri M, Carpentier M-C, Lasserre E, Cooke R, Rosa Farias D da, da Maia LC, dos Santos RS, Nyberg KG, McNally KL, Mauleon R, Alexandrov N, Schmutz J, Flowers D, Fan C, Weigel D, Jena KK, Wicker T, Chen M, Han B, Henry R, Hsing YC, Kurata N, de Oliveira AC, Panaud O, Jackson SA, Machado CA, Sanderson MJ, Long M, Ware D, Wing RA. 2018. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. *Nat Genet* **50**:285–296. doi:10.1038/s41588-018-0040-0

Sugihara Y, Abe Y, Takagi H, Abe A, Shimizu M, Ito K, Kanzaki E, Oikawa K, Kourelis J, Langner T, Win J, Białas A, Lüdke D, Contreras MP, Chuma I, Saitoh H, Kobayashi M, Zheng S, Tosa Y, Banfield MJ, Kamoun S, Terauchi R, Fujisaki K. 2023. Disentangling the complex gene interaction networks between rice and the blast fungus identifies a new pathogen effector. *PLOS Biol* **21**:e3001945. doi:10.1371/journal.pbio.3001945

Tamborski J, Seong K, Liu F, Staskawicz BJ, Krasileva KV. 2023. Altering Specificity and Autoactivity of Plant Immune Receptors Sr33 and Sr50 Via a Rational Engineering Approach. *Mol Plant-Microbe Interactions®* **36**:434–446. doi:10.1094/MPMI-07-22-0154-R

Toghani A, Kamoun S. 2024. Functional annotation of 180 RefSeq reference plant proteomes reveals a dataset of 113,684 NLR proteins. doi:10.5281/zenodo.13627395

Wang Jizong, Hu M, Wang Jia, Qi J, Han Z, Wang G, Qi Y, Wang H-W, Zhou J-M, Chai J. 2019. Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science* **364**:eaav5870. doi:10.1126/science.aav5870

Zhao Y-B, Liu M-X, Chen T-T, Ma X, Li Z-K, Zheng Z, Zheng S-R, Chen L, Li Y-Z, Tang L-R, Chen Q, Wang P, Ouyang S. 2022. Pathogen effector AvrSr35 triggers Sr35 resistosome assembly via a direct recognition mechanism. *Sci Adv* **8**:eabq5108. doi:10.1126/sciadv.abq5108

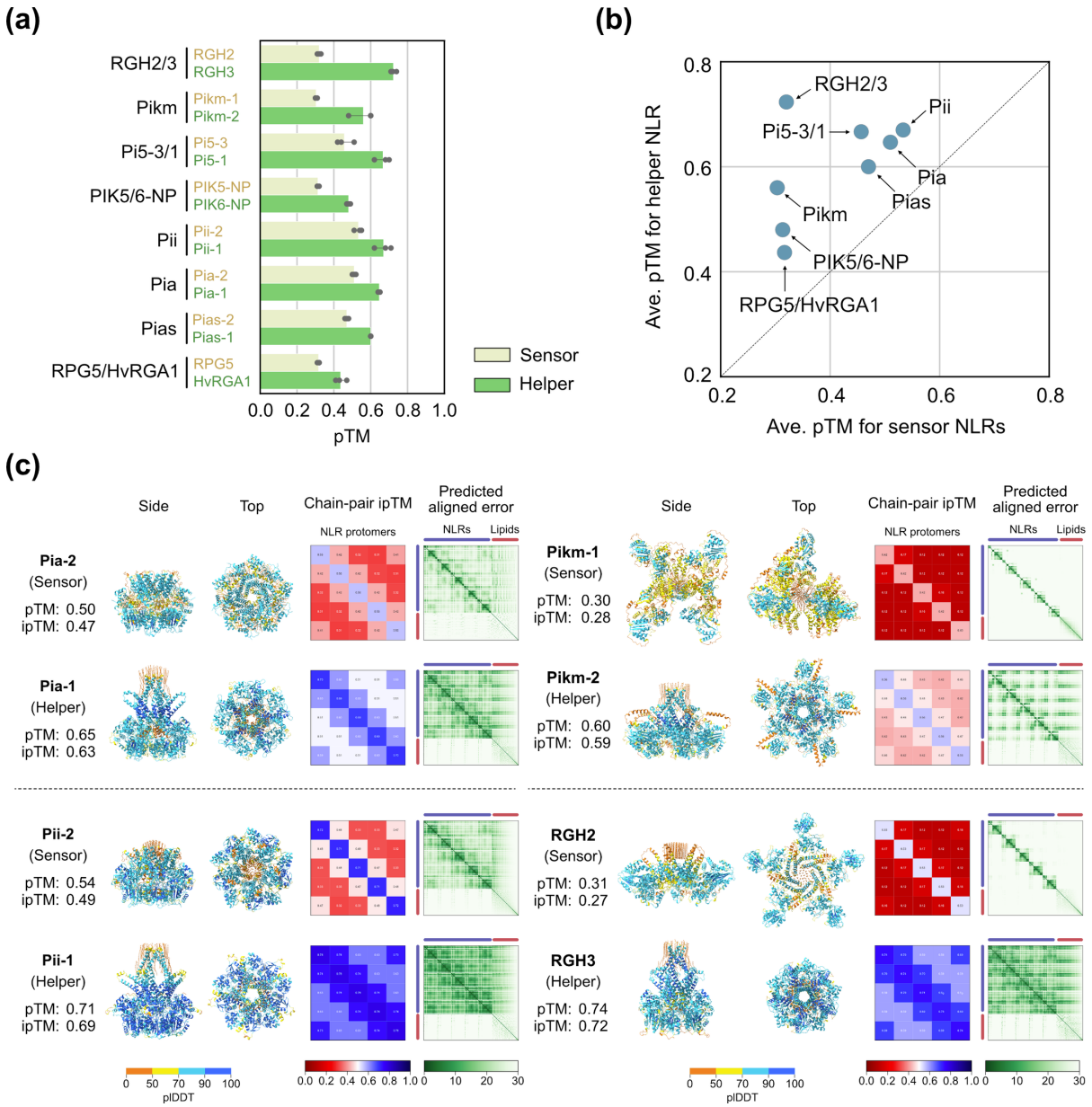


Fig. 1. Helper NLRs produce higher AlphaFold 3 pTM scores than their paired sensors. a) Bar plot comparing sensor and helper pTM scores in pentameric AlphaFold 3 predictions. The amino acid sequences of the oligomerizing domains of the NLR proteins, from the N-terminus to the end of the NB-ARC domain, were used for the prediction. The pentameric structures were modelled with 50 oleic acids using three different seed values. b) Scatter plot comparing sensor and helper pTM scores in pentameric AlphaFold 3 predictions. The resulting pTM scores were averaged across three seed values for each sensor and helper NLR. c) Pentameric AlphaFold 3 structures of the four representative NLR pairs. The structures predicted using seed value 1 were visualized using ChimeraX (Meng et al., 2023).

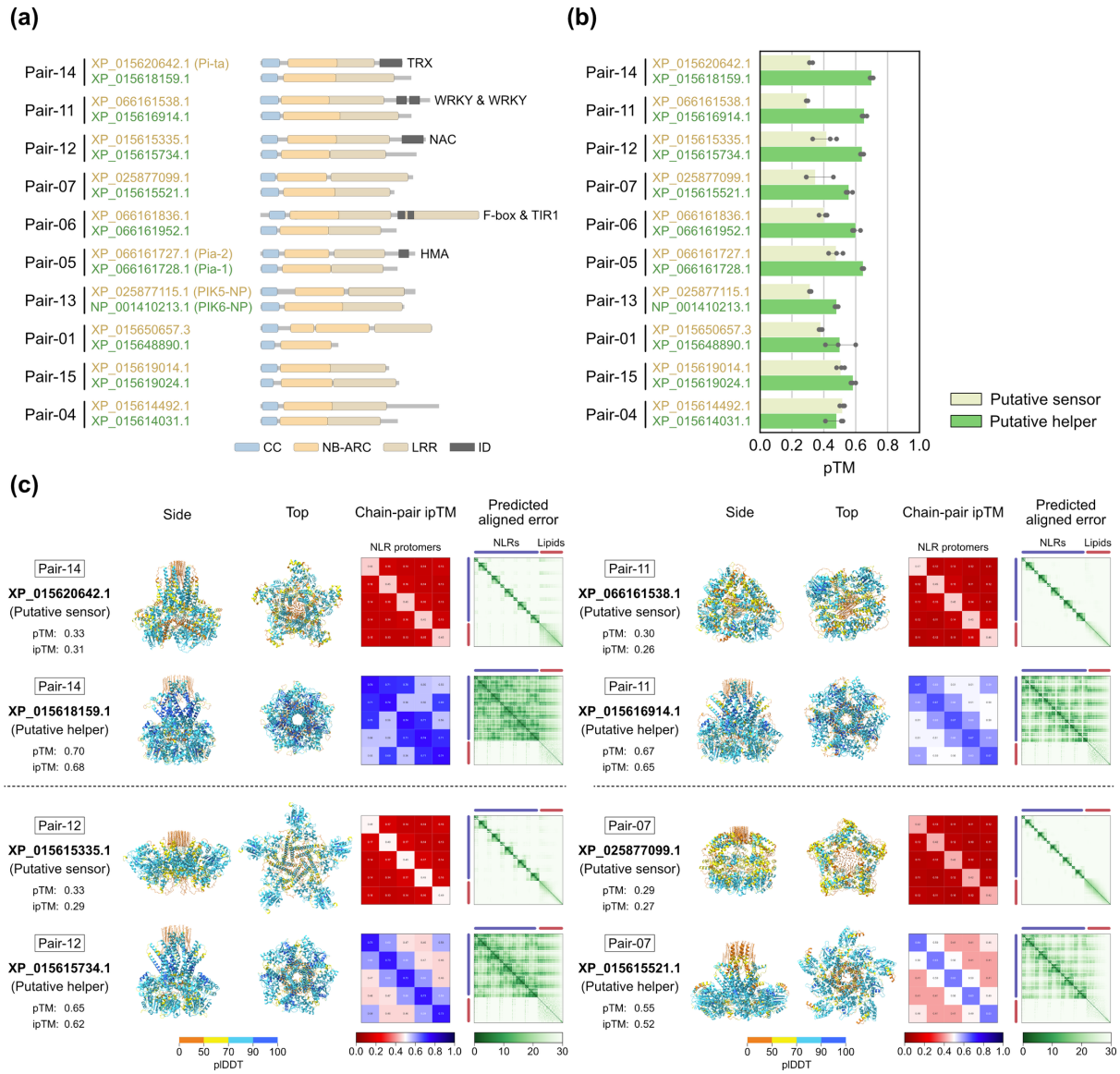


Fig. 2. AF3 can classify paired NLRs even when the putative sensors do not carry an integrated domain annotation. a) Domain architectures of ten rice NLR pairs described by Stein et al., 2018. The domains were annotated with NLRtracker (Kourelis et al., 2021) and visualized with refplantnlR (https://github.com/JKourelis/refplantnlR). b) Bar plot comparing putative sensor and helper pTM scores in pentameric AlphaFold 3 predictions. The amino acid sequences of the oligomerizing domains of the NLR proteins, from the N-terminus to the end of the NB-ARC domain, were used for the prediction. The pentameric structures were modelled with 50 oleic acids using three different seed values. Putative sensors and helpers were assigned based on the average pTM scores of pentameric and hexameric structures, with a putative sensor having the lower average score and a putative helper having the higher average score. c) Pentameric AlphaFold 3 structures of the four representative NLR pairs described by Stein et al., 2018. The structures predicted using seed value 1 were visualized using ChimeraX (Meng et al., 2023).

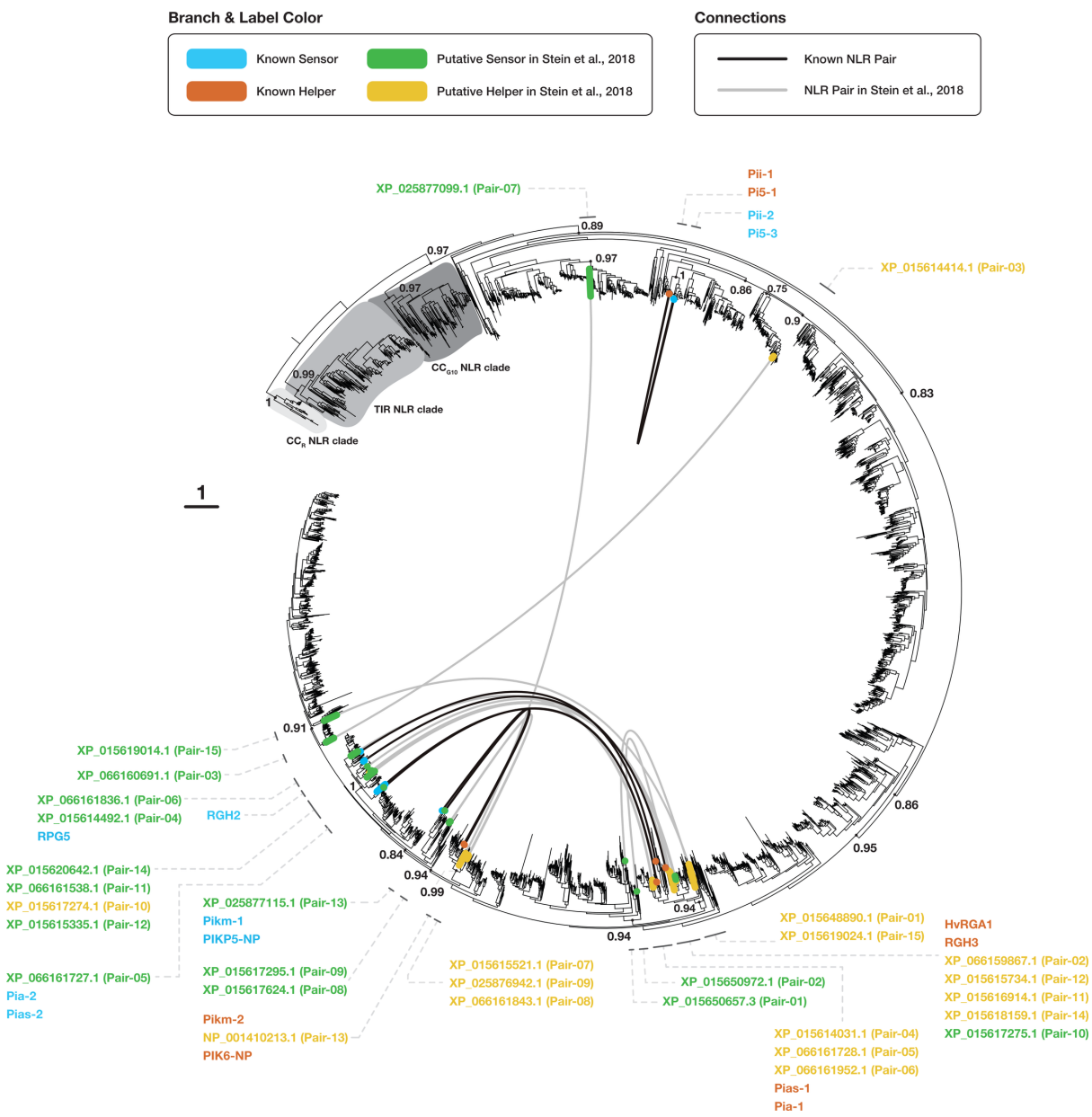
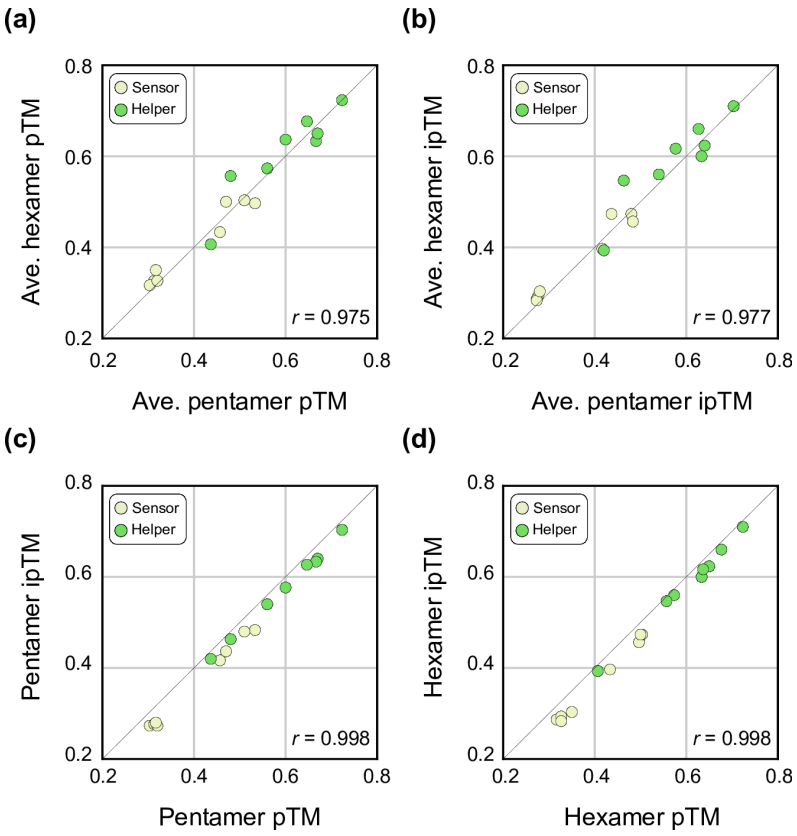


Fig. S1. Phylogenetic tree of monocot paired NLRs with RefPlantNLR and 4,936 NLR proteins from 13 RefSeq proteomes, including two dicot species and 11 monocot species obtained from Toghani et al., 2024. The tree was built using the NB-ARC domains of the eight known NLR pairs summarized in **Table S1**, RefPlantNLR, and 4,936 NLR proteins from the NLRtracker output of 13 RefSeq proteomes, including two dicot species (*Arabidopsis thaliana* and *Solanum lycopersicum*) and 11 monocot species (*Zea mays*, *Triticum aestivum*, *Setaria viridis*, *Phragmites australis*, *Phoenix dactylifera*, *Oryza sativa*, *Musa acuminata*, *Lolium perenne*, *Hordeum vulgare* subsp. *vulgare*, *Brachypodium distachyon*, and *Asparagus officinalis*) (**Data S1, S2, S3, and S4**) (Toghani and Kamoun, 2024). The paired NLRs of rice cultivar Nipponbare (GCF_034140825.1), described by Stein et al., 2018, are included among 13 RefSeq proteomes. The genetic linkage between NLR pairs is indicated by a connection between linked nodes. Known sensor and helper NLR nodes are colored blue and orange, respectively, while putative sensor and helper nodes are colored green and yellow. Numbers next to the tree nodes indicate bootstrap values. The tree is rooted at the CC$_R$-NLR clade. AlphaFold 3 was not performed for Pair-02, Pair-03, Pair-08, Pair-09, and Pair-10 because they do not carry a full N-terminal CC domain.



411
412
413 **Fig. S2. Correlations between pentameric and hexameric AlphaFold 3 confidence scores and between**
414 **pTM and ipTM scores for previously reported NLR pairs.** a) Correlation between the average pentameric and
415 hexameric pTM scores. b) Correlation between the average pentameric and hexameric ipTM scores. c)
416 Correlation between the pentameric pTM and ipTM scores. d) Correlation between the hexameric pTM and ipTM
417 scores. The amino acid sequences of the oligomerizing domains of the NLR proteins, from the N-terminus to the
418 end of the NB-ARC domain, were used for the prediction. The pentameric and hexameric structures were
419 modelled with 50 oleic acids using three different seed values. The resulting pTM or ipTM scores were averaged
420 across three seed values for each NLR in Fig. S2a and S2b. The Pearson rank correlation coefficients (*r*) were
421 calculated with the SciPy library in Python.

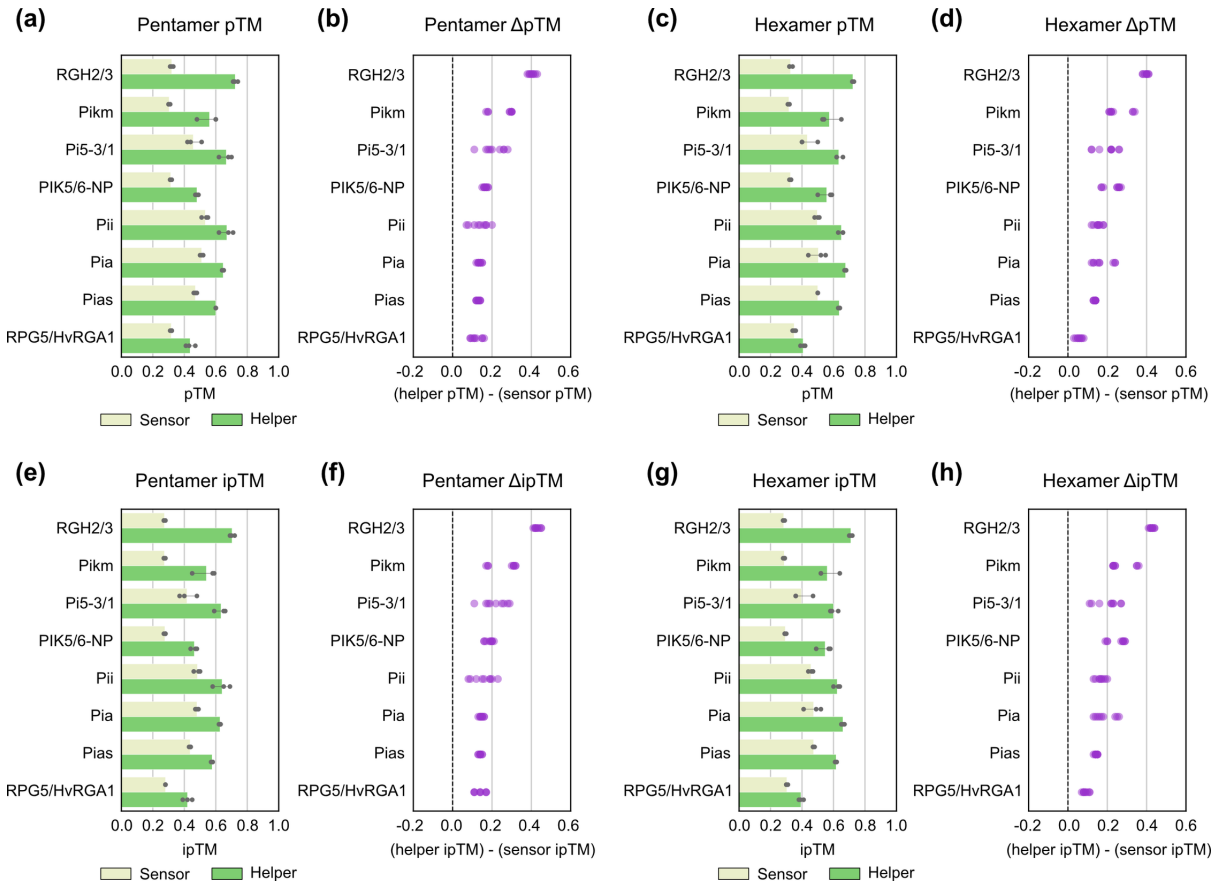


Fig. S3. Comparisons of sensor and helper AlphaFold 3 scores in pentameric and hexameric configurations for previously reported NLR pairs. a-d) pTM score comparisons for pentamers (a,b) and hexamers (c,d). e-h) ipTM score comparisons for pentamers (e,f) and hexamers (g,h). The a, c, e, and g panels show bar plots, and the b, d, f, and h panels display score differences (helper minus sensor). The amino acid sequences of the oligomerizing domains of the NLR proteins, from the N-terminus to the end of the NB-ARC domain, were used for the prediction. The pentameric and hexameric structures were modelled with 50 oleic acids using three different seed values. Subtractions were performed for all possible pairs of sensor and helper scores.

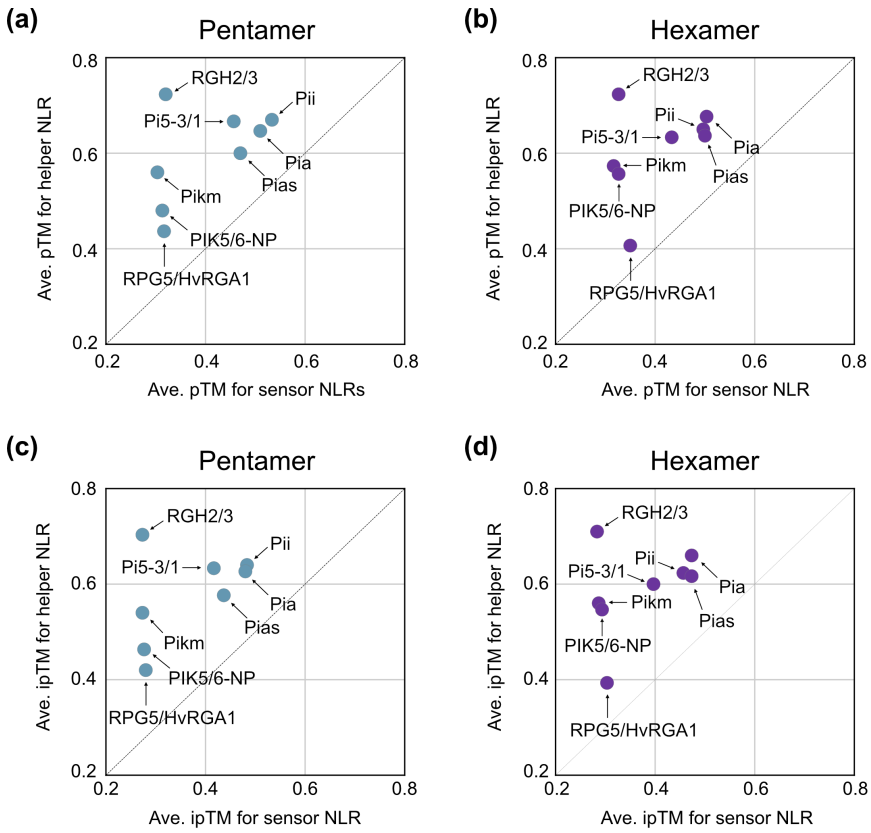


Fig. S4. Scatter plots comparing average sensor and helper scores in AlphaFold 3 predictions. a) Scatter plot comparing average sensor and helper pTM scores in pentamer predictions. b) Scatter plot comparing average sensor and helper pTM scores in hexamer predictions. c) Scatter plot comparing average sensor and helper ipTM scores in pentamer predictions. d) Scatter plot comparing average sensor and helper ipTM scores in hexamer predictions. The amino acid sequences of the oligomerizing domains of the NLR proteins, from the N-terminus to the end of the NB-ARC domain, were used for the prediction. The pentameric and hexameric structures were modelled with 50 oleic acids using three different seed values. The resulting pTM or ipTM scores were averaged across three seed values for each sensor and helper NLR.

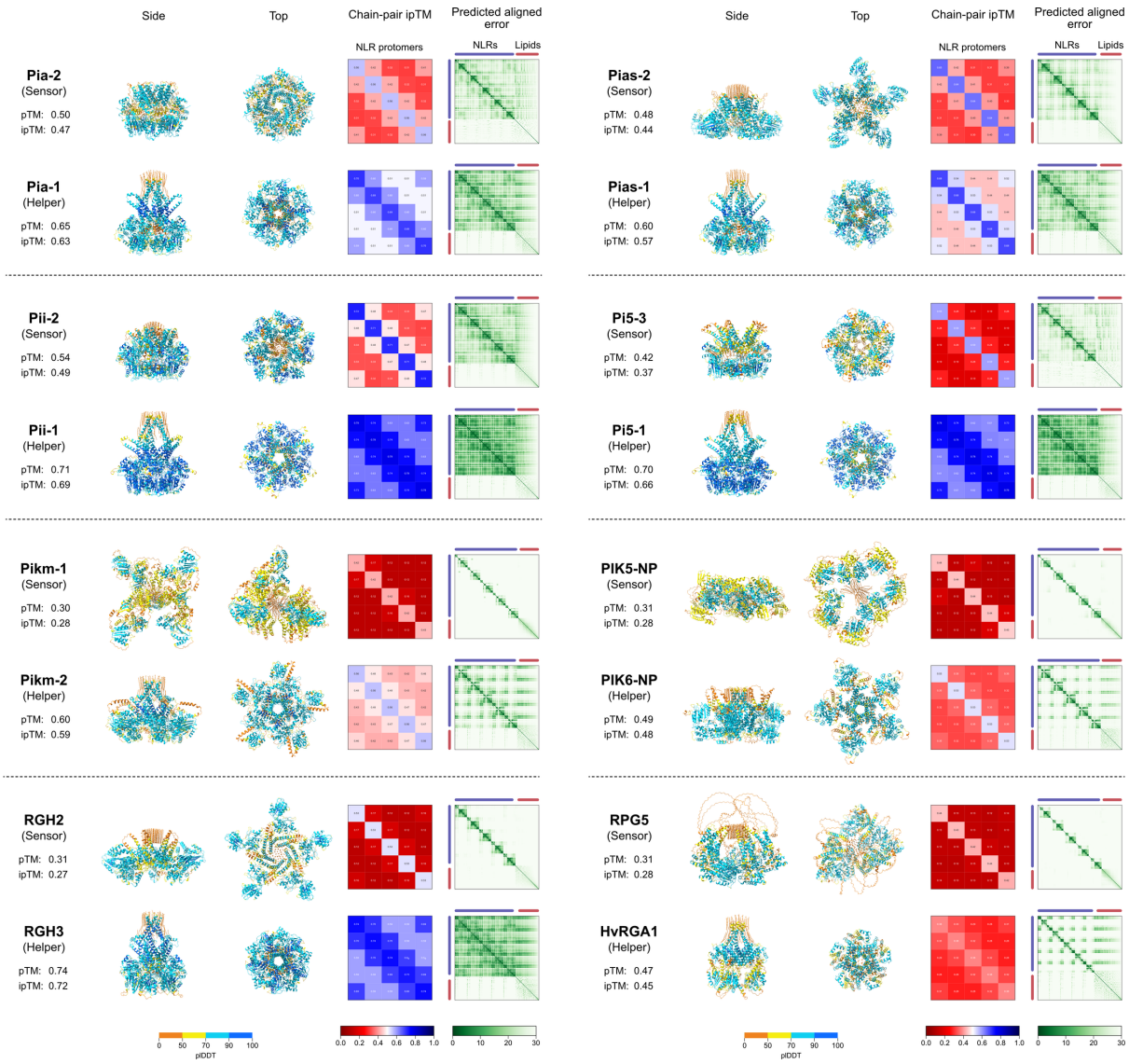


Fig. S5. Pentameric AlphaFold 3 predictions for previously reported NLR pairs. The amino acid sequences of the oligomerizing domains of the NLR proteins, from the N-terminus to the end of the NB-ARC domain, were used for the prediction. The pentameric structures were modelled with 50 oleic acids using seed value 1. The predicted structures were visualized using ChimeraX (Meng et al., 2023).

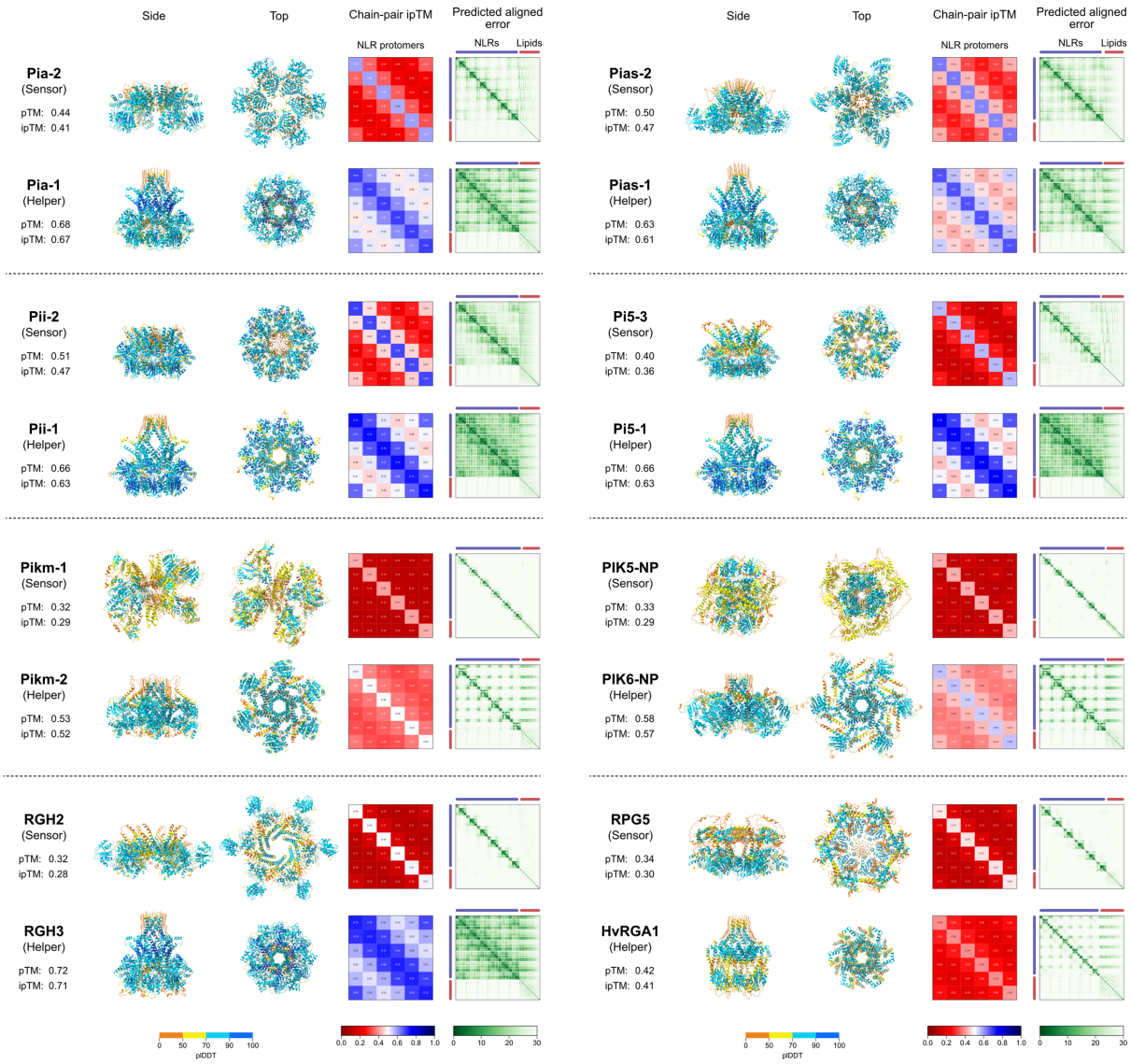


Fig. S6. Hexameric AlphaFold 3 predictions for previously reported NLR pairs. The amino acid sequences of the oligomerizing domains of the NLR proteins, from the N-terminus to the end of the NB-ARC domain, were used for the prediction. The hexameric structures were modelled with 50 oleic acids using seed value 1. The predicted structures were visualized using ChimeraX (Meng et al., 2023).

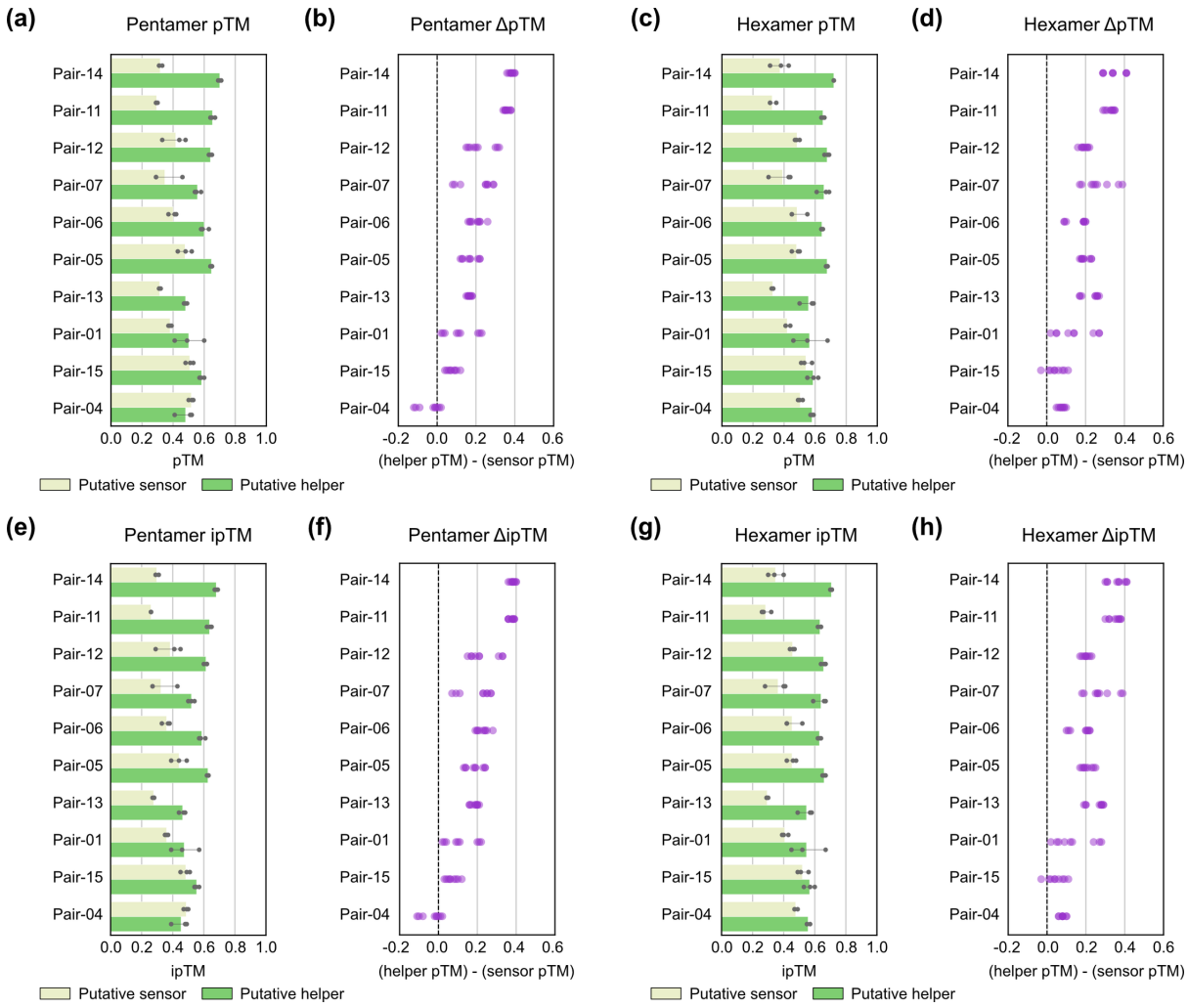


Fig. S7. Comparisons of putative sensor and helper AlphaFold 3 scores in pentameric and hexameric configurations for NLR pairs described by Stein et al., 2018. a-d) pTM score comparisons for pentamers (a,b) and hexamers (c,d). e-h) ipTM score comparisons for pentamers (e,f) and hexamers (g,h). The a, c, e, and g panels show bar plots, and the b, d, f, and h panels display score differences (putative helper minus putative sensor). The pentameric and hexameric structures were modelled with 50 oleic acids using three different seed values. Putative sensors and helpers were assigned based on the average pTM scores of pentameric and hexameric structures, with a putative sensor having the lower average score and a putative helper having the higher average score. Subtractions were performed for all possible pairs of putative sensor and helper scores.

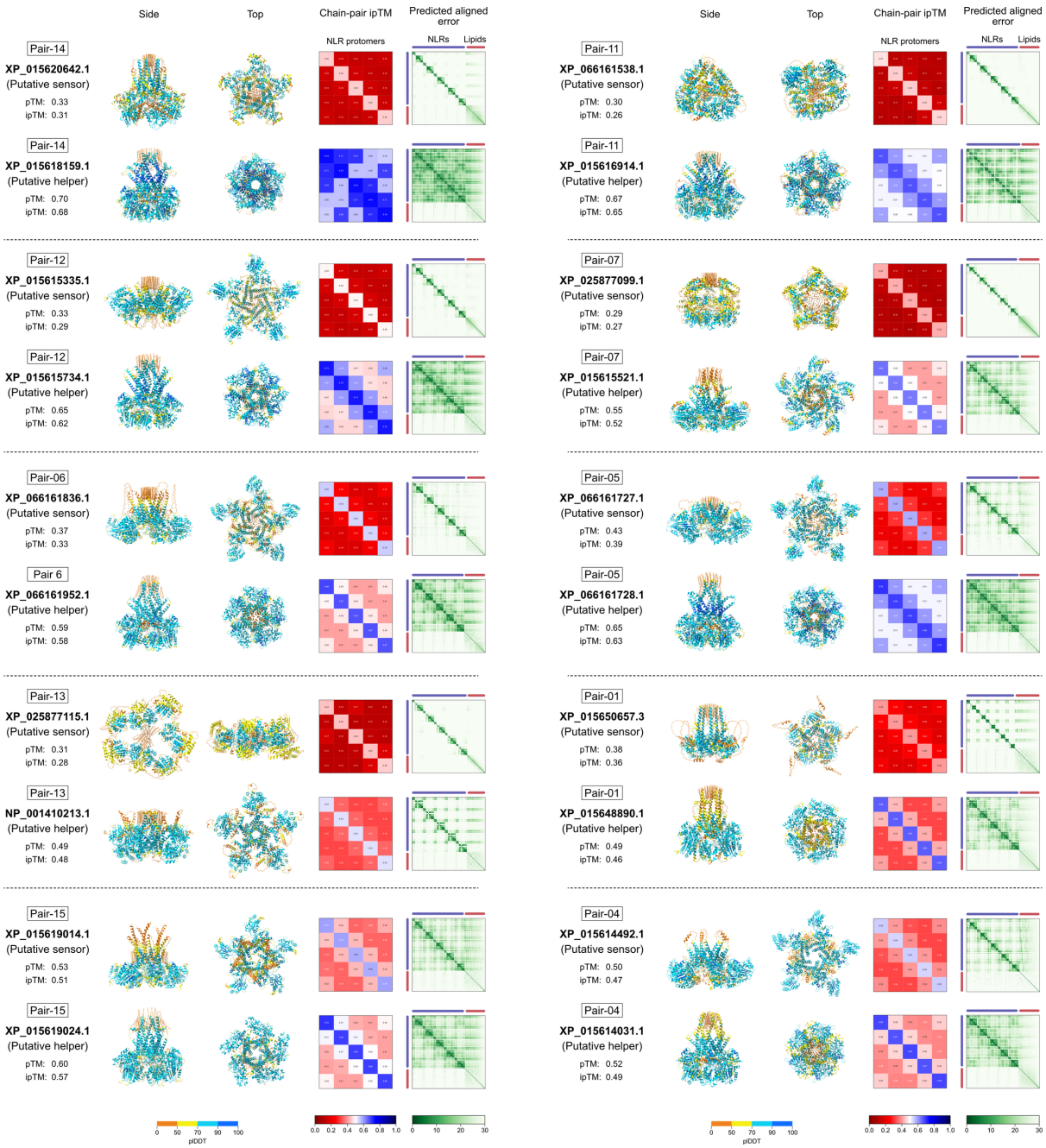


Fig. S8. Pentameric AlphaFold 3 predictions for NLR pairs described by Stein et al., 2018. The amino acid sequences of the oligomerizing domains of the NLR proteins, from the N-terminus to the end of the NB-ARC domain, were used for the prediction. The pentameric structures were modelled with 50 oleic acids using seed value 1. The predicted structures were visualized using ChimeraX (Meng et al., 2023).

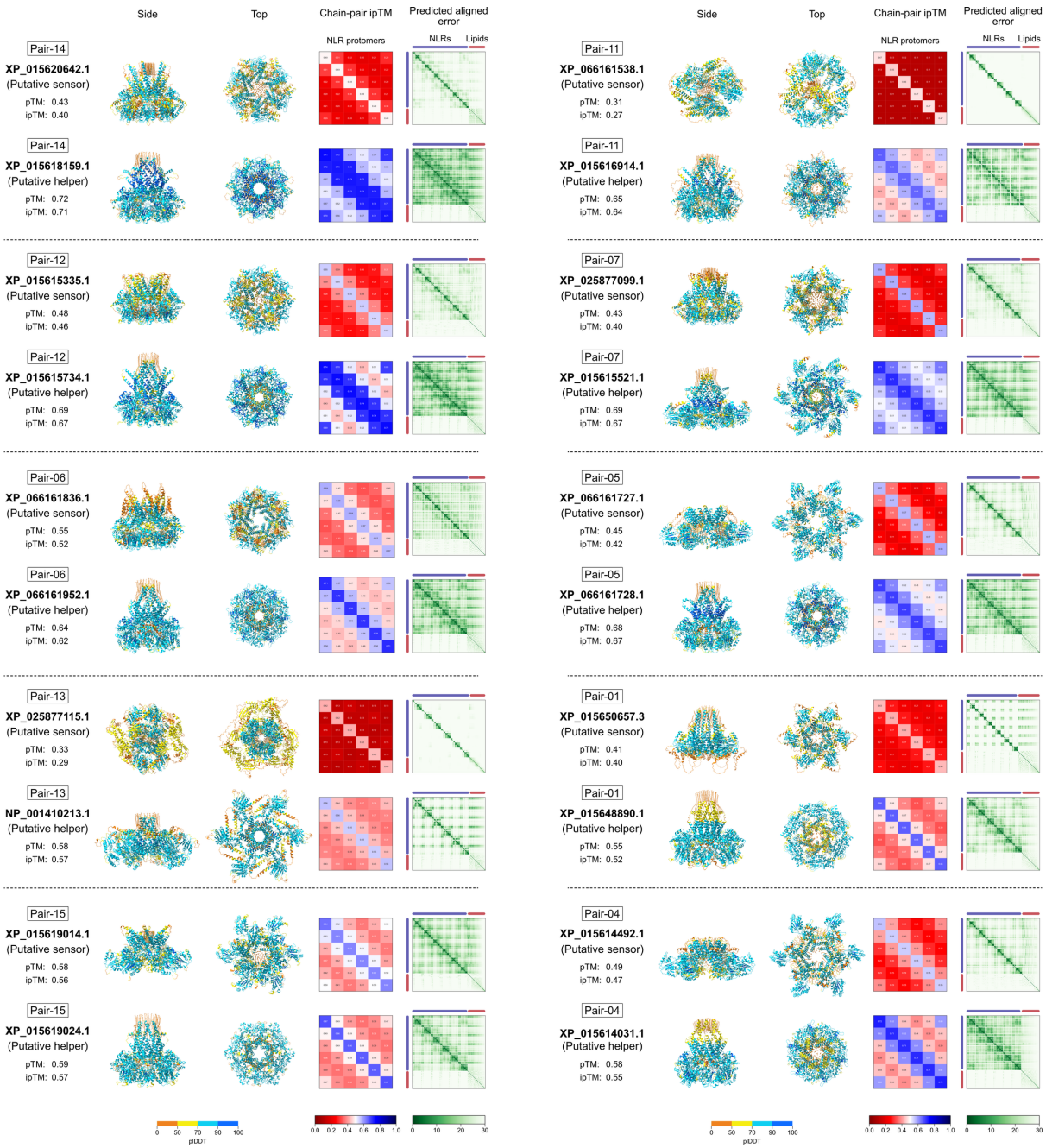


Fig. S9. Hexameric AlphaFold 3 predictions for NLR pairs described by Stein et al., 2018. The amino acid sequences of the oligomerizing domains of the NLR proteins, from the N-terminus to the end of the NB-ARC domain, were used for the prediction. The hexameric structures were modelled with 50 oleic acids using seed value 1. The predicted structures were visualized with ChimeraX (Meng et al., 2023).

Table S1. List of previously reported NLR pairs. All sequences were derived from either RefPlantNLR (Kourelis et al., 2021) or NCBI. Sensor and helper NLRs are defined based on the presence or absence of an integrated domain, respectively. Domain architectures were annotated using NLRtracker (Kourelis et al., 2021). For domain architecture, "C," "N," "L," and "O" represent the CC, NB-ARC, LRR, and other integrated domains, respectively. Note that InterProScan did not annotate an integrated domain of PIK5-NP as previously reported (Kourelis et al., 2021). Therefore, we manually replaced the domain architecture of PIK5-NP from "CNL" to "CONL" as it contains the HMA domain between the NB-ARC and LRR (Białas et al., 2021).

Table S2. Summary of AlphaFold 3 predictions for previously reported NLR pairs.

Table S3. HMM scores of MADA motifs in previously reported NLR pairs. The presence or absence of the MADA or MADA-like motif was analyzed using the HMM model in Adachi et al., 2019a. MADA, MADA-like, and no MADA motifs were classified based on HMM scores (> 10, < 10, and NA, respectively).

Table S4. List of rice NLR pairs described by Stein et al., 2018. The NLR pairs were identified from the rice cultivar Nipponbare genome annotation (Stein et al., 2018). Based on Supplementary Data 6 in Stein et al., 2018, the NLR pairs that i) are genetically linked in head-to-head orientations; ii) belong to distinct phylogenetic clades were extracted. Domain architectures were annotated using NLRtracker (Kourelis et al., 2021). Columns from Supplementary Data 6 in Stein et al., 2018 and those added in this study are highlighted in yellow and green, respectively.

Table S5. BLASTP results using NLRs from Stein et al., 2018 as queries against those in the NCBI RefSeq annotation as subjects. DIAMOND BLASTP (Buchfink et al., 2021) was used to identify corresponding sequences between the datasets. NLRs from Stein et al., 2018 and those in the NCBI RefSeq annotation of the rice cultivar Nipponbare (GCF_034140825.1) served as the query and subject sets, respectively.

Table S6. Summary of NLRtracker outputs for rice NLR pairs from the NCBI RefSeq annotation of Nipponbare (GCF_034140825.1), corresponding to those described by Stein et al., 2018.
Only NLR pairs with a full N-terminal CC domain in both NLRs were processed for AlphaFold 3 predictions.

Table S7. Summary of AlphaFold 3 predictions for rice NLR pairs from the NCBI RefSeq annotation of Nipponbare (GCF_034140825.1), corresponding to those described by Stein et al., 2018.
Putative sensors and helpers were assigned based on the average pTM scores of pentameric and hexameric structures, with a putative sensor having the lower average score and a putative helper having the higher average score.

Data S1. Taxonomic and genome metadata of species used for phylogenetic analysis.

Data S2. Sequence and metadata of full-length and NB-ARC domains of NLR sequences used for phylogenetic analysis.

Data S3. Alignment of 5,509 NB-ARC sequences of NLRs used in Fig. S1.

Data S4. Phylogenetic tree of 5,509 NB-ARC sequences of NLRs used in Fig. S1.