Qeios

Peer Review

Review of: "Can AI Modelling of Protein Structures Distinguish Between Sensor and Helper NLR Immune Receptors?"

Sivakumar Prasanth Kumar[1]

1. St. Jude Children's Research Hospital, Memphis, United States

The authors used AlphaFold3 (AF3), an artificial intelligence-based protein structure prediction approach, to classify paired NLR proteins into sensor or helper categories based on the ability to predict/form oligomeric structure. Analyzing known paired sensor-helper NLRs, known homologs, and extending to pairs that contain non-canonical domains is an interesting way to understand why certain NLRs can form high-order structures (e.g., funnel-like) essential for pathogen detection, possibly classifying them as putative helper NLRs. Though the study is novel with better applications towards structure-based classification of NLRs (helper/sensor) in plants, there are several key aspects that have been overlooked to appreciate such an AF3 role in functional protein classification based purely on the oligomeric structure and AF3 metric (pTM; template modeling score).

1. The authors have not provided the domain architecture of plant NLRs and indicated which sets of domains/proteins are experimentally characterized to form or not form high-order structures, either by structure identification (cryo-EM) or biochemical validations. This would provide a basic understanding of which domains tend to form high-order structures. Even Table S1 lists whether a protein in the helper-sensor pair is experimentally validated, but does not provide the PDB entries. Readers could benefit from a figure containing the domain architecture of characterized plant NLRs and some of their experimentally identified PDB structures.

2. Which domains of the helper-sensor pair are structurally solved to monomeric and high-order forms? It is plausible that AF3, trained on the Protein Data Bank data containing plant NLRs, might predict structures with good confidence, but not for unseen domain compositions. Did the authors perform structure prediction for distantly related protein pairs? How much is the dataset's sequence and structural similarity to PDB data?

doi.org/10.32388/C6DPQY

3. Which structurally computed pairs are similar to experimentally characterized pairs or in isolation in terms of RMSD, congruity between high/low confidence regions with abundance, and lack of electron density?

4. Did the authors perform randomization in the domain compositions of sequences to identify whether AF3 was hallucinating structures with/without IDs?

5. The study currently focuses on coiled-coil NLRs in very few plants, rice and barley. Though it is very difficult to predict helper-sensor structure pairs for uncharacterized plants of agricultural importance (e.g., paddy, wheat, cereals) using AF3, the authors are recommended to identify distant homologs through sequence-based searches and develop putative models to check whether the same trend of score and oligomeric forms is still valid for such cases.

6. The reviewer could not comprehend the usage and impact of oleic acids in determining the oligomeric form of helper-sensor protein pairs. Though oleic acid mimics the lipid environment, how oleic acid facilitates the proper folding of oligomeric complexes (e.g., funnel-shaped helper NLRs) is lacking.

7. Figure 1 shows a cartoon representation of helper-sensor pairs. It is visually appealing that most of the regions in the complexes have plDDT scores in the range of 50 to 70 (the yellow region in the plDDT wand). Why are there low-confidence regions in the complexes? Is it due to a lack of sequence similarity with known structures or intrinsically disordered regions? Such low-confidence regions might directly influence the pTM score and hence introduce inherent bias in considering the pTM score for classifying NLRs into either helper or sensor components.

8. Why did the authors bring the HMM score from the MADA motif-based HMM model to classify helper/sensor components just by the presence of the MADA motif in their dataset sequences? Why is it being compared with AF3, and why could the authors not classify most of their dataset proteins with the HMM score, except for Pikm and PIK5/6-NP? These are not explained.

The manuscript has several instances of grammatical inconsistencies, improper sentence construction, and issues with text flow. For example, pentameric or hexameric pore-like complexes are introduced, then the funnel-shaped structure is discussed subsequently, and NB-ARC is introduced in the first paragraph of the results. It is currently difficult for a broader audience to understand the foundations of plant NLRs, domain compositions, high-order structures, plants that

have been studied for helper-sensor pairs, experimental data that are available over the internet, and so on.

Declarations

Potential competing interests: No potential competing interests to declare.