Use and Limitations of AlphaFold2 in the Structural Modeling of E3 Ubiquitin Ligase Tom1/HUWE1

Jennifer Madrigal[1]
[1]University of Utah
j.madrigal@utah.edu

Tom1 is a 374 kDa E3 ubiquitin ligase found in yeast, and is highly conserved from yeast to humans. Several forms of cancer and intellectual disabilities have been associated with disruption of the human homolog HUWE1, and this enzyme has become a potential therapeutic target in the treatment of cancer. Although it has been found to play a role in several discrete cellular processes, the major enzymatic preference of the protein appears to be directed at RNA-binding proteins which are removed from their RNA binding partners. The large n-terminal alpha-solenoid upstream of the catalytic HECT domain gives Tom1/HUWE1 it's substrate-specific catalytic activity, but it remains unclear how substrates are selected and how catalytic activity is regulated. Additionally, the development of inhibitors to be used in cancer treatments has been limited without an understanding of the structure of the regulatory region of Tom1/HUWE1 which lies N-terminal to the HECT domain. Although both the yeast and human homologs were too large to be included in the first AlphaFold Protein Structure Database release, an AlphaFold2 prediction of the zebrafish HUWE1 has been deposited. This prediction includes many large unstructured loops with low-confidence scores, and even the well-characterized c-terminal HECT domain is incorrectly predicted, instead appearing as a series of low-confidence unstructured loops. However, the general form of the HUWE1 body is predicted with much higher confidence and indicates a structure in which a succession of ARM-like repeats connects the N-terminal and C-terminal regions in an alpha-solenoid fashion.

Using single-particle cryo-EM, we reconstructed a map of apo-Tom1/HUWE1 in its most well-ordered closed-conformation form to 3.0 Å overall, allowing the confident modeling of most of the well-ordered alpha-solenoid portion of the molecule. However, many domains protruding between the ARM-like repeats of the solenoid are more flexibly tethered, increasing the difficulty of modeling residues in-register with high confidence. AlphaFold2 was used to predict small regions of the molecule at a time (100-200 residues in length) which models gave surprising agreement with the map compared to the poorly modeled predictions of the overall HUWE1 structure. These predictions both validated residues already modeled and assisted in confidently correcting out-of-register residues in other areas of the molecule.

The secondary and tertiary structure of Tom1/HUWE1 has largely eluded researchers in part due to flexibility of the molecule, but also due to a lack of sequence homology resulting in poor secondary structure prediction. Lack of sequence homology is likely a large part of the reason for the poor AlphaFold prediction, but there are other factors to consider as well. Improvements between the first AlphaFold release and AlphaFold2 were in part designed to overcome the limitations of bias that occurred from processing structural predictions in sets of residues in sequence-proximity to each other, yet the neural network of AlphaFold2 was still unable to accurately predict the N-C terminal interactions in a full-length model. However, it is likely that a combination of factors, including the greater number of surface-exposed residues in Tom1/HUWE1, low sequence homology relating to known structures, and possible other confounding factors, which can largely explain the inaccurate predicted intra-molecular interactions leading to an inability to accurately predict even the previously characterized and highly homologous HECT domain, which limitations were partially resolved by piecemeal domain-by-domain structural predictions. This case study of the unusual structure of Tom1/HUWE1 and comparison of the experimentally derived structure with predicted structures may provide insight into considerations for subsequent improvements of structural prediction software such as AlphaFold.