Development of an algorithm to evaluate 3D protein structure prediction similarities due to mutations

Jongkeun Park, Seung Pil Jeong, Downwan Hong

Department of Medical Informatics, Graduate School, The Catholic University of Korea, Seoul, Korea

(bellpark@catholic.ac.kr)

Recently, with the development of artificial intelligence-based protein 3D structure prediction technology, it has become easier to identify protein structures and analyze their folding, docking, and pockets. In that process, researchers use a molecular visualization system to directly inspect the protein structure based on PDB files that are results of protein structure prediction. However, to analyze large amounts of protein structural information, an automated algorithm that evaluates protein structure similarity is required.

We developed the following algorithm to compare the variants against a reference. The input is a PDB file generated by AlphaFold2. The pLDDT score, which is the result of AlphaFold2, selects a region that shows a difference from the reference. For the selected region, the peptide torsion angles phi, psi, and omega were compared and analyzed. Through Root Mean Square Deviation (RMSD), the similarity in the distance between atoms within the peptide torsion angles was assessed. In addition, we compared the difference in area due to mutations against the reference by examining the area between amino groups (-NH2) and carboxyl groups (-COOH). A mathematical model incorporating the slope and area between amino acid sequences automatically evaluated the similarity between protein structures.

We implemented an automated and highly accurate algorithm, and the results of this research will be usable for candidate drug development and drug repositioning.