Rethinking Drug Discovery: New Algorithms for Virtual Drug Screening

Punjabi, Amol

On average, pharmaceutical companies spend $2.6 billion and twelve years to develop a single drug. Most small molecule drugs are discovered by high-throughput robotic screening, an expensive and time-consuming process for discovering ligands (drug candidates) that bind to drug targets. Virtual drug screening docks ligands onto a cellular target structure and predicts their binding affinities. However, only 2% of ligands with strong predicted affinities are true binders in vitro. Two algorithms were developed in the language R to improve binding affinity prediction accuracy. The first algorithm determined the most frequent structure of a protein from a molecular dynamics simulation and used this for ligand docking. This was achieved by minimizing the biased Root Mean Squared Deviation between molecular dynamics protein conformations and by optimizing the program for parallel computing. The second algorithm built a machine learning model from public mutagenesis data which discerned spatial and electrostatic characteristics of strong binding. This was achieved by analyzing Spearman's rank correlation coefficients to determine characteristics of substrate-protein binding and designing a statistical similarity test resistant to outliers to identify true binders. These algorithms were validated on the target APOBEC3A, an enzyme that induces cancer-causing mutations. The affinities of the top eight ligands to APOBEC3A predicted by the algorithms and by standard docking were compared in vitro. While the correlation of binding affinities to standard docking predictions was 0.26, the correlation to algorithmic predictions was 0.80. These novel algorithms could streamline the time and cost of discovering future therapeutics for life-threatening diseases.

Awards Won:

First Award of $5,000

Intel ISEF Best of Category Award of $5,000