A Novel Computational Generative Model That Identifies Ligands: A Potential Way to Reduce Cost & Time in Identifying Promising Drug Candidates

Ende, Chase (School: Harrison High School)

Drug discovery is intricate and costly. Current methods in de novo drug design rely on undiversified and scarce datasets due to the impossibility of recording all intermolecular interactions on the atomic level between ligands and receptors. This limitation renders these models inaccurate by leading to flawed or skewed binding predictions. Therefore, it is pressing to create a novel computational drug design method that can provide accurate binding affinity data for evaluating the effectiveness of developed drugs. This study aimed to address the limitations of current de novo drug design methods, which cannot quantitatively evaluate interactions necessary for determining drug efficacy. The hypothesis was that an unsupervised learning model, which does not require a large amount of initial data, can generate drug molecules with stronger binding affinities to a target receptor than its natural ligands. Phase 1 (2023) involved successfully mapping out the cavities in a selected receptor molecule using custom Raycasting. Phase 2 (2024) added a genetic algorithm—a type of unsupervised learning model that uses a crossover function to converge to an optimal solution—to find the amino acid chain with the highest binding affinity, and created a neural network to rapidly generate ligands. Results showed that each computationally generated ligand had binding affinity values consistently surpassing the accepted value of the natural ligand's binding affinity. This novel, unsupervised computational drug design model demonstrates the potential to create drugs with stronger binding affinities to specific target receptors, which is indicative of enhanced drug efficacy, and could potentially reduce costs and enhance the efficiency of early-stage drug development.

Awards Won: