DiPPI: Drugs in Protein-Protein Interaction Interfaces

Fatma	Cankara	fcankara20@ku.edu.tr	Turkey	Koç University
Simge	Şenyüz	ssenyuz19@ku.edu.tr	Turkey	Koç University
Özlem	Keskin	okeskin@ku.edu.tr	Turkey	Koç University
Attila	Gürsoy	agursoy@ku.edu.tr	Turkey	Koç University

Protein-protein interactions (PPIs) modulation has been associated with various diseases. Proteins interact through their interfaces and interface characterization is important to understand the underlying mechanisms of the interaction. There has been an increasing interest in PPIs as drug targets, as analysis of binding regions and critical contacts may reveal more information about these mechanisms. Therefore, investigating the properties of the interactions is critical for the identification of drug targets, action mechanisms, and alternative pathways for drug repurposing. Here, we present a drug target dataset that can be utilized for drug repurposing studies with a focus on interface-bound drugs. Our dataset has two modules for drugs and interfaces. On the interface side, we extracted several properties of interfaces such as amino acid propensities, hotspots, evolutionary conservation of drug binding amino acids, and post-translational modifications of its residues. The interfaces are extracted from PDB and clustered based on their structural properties in our previous study. This broad characterization helps to identify similar interfaces with a focus on both physicochemical properties and structural similarity. On the drug side, we curated drug-like small molecules and FDA-approved drugs from various databases and created a subset that contains drugs bound to these interfaces. The drugs are clustered based on their molecular fingerprints to limit the search for an alternative drug to a smaller space. Drug properties such as following Lipinski's rule and various molecular descriptors are also analyzed to further guide the selection of drug molecules. In total, we created a dataset consisting of 11,011 drug-like small molecules of which 343 are FDA-approved drugs. Filtered drug-like small molecules are found in 335,648 interfaces. 19,994 interfaces have an FDA-approved drug bound to them. Our highly redundant dataset provides users with an easy-to-follow scheme for drug repurposing studies through its well-curated and clustered interface and drug data.