A New Method for Simulation-Aided Drug Discovery Using Machine Learning Algorithms

Ram, Parvathi

Time, effort, cost involved in discovering candidate peptides which can act as potential drugs against the virulence factors of disease-causing pathogens, are huge. Numerous docking trials are needed to determine the binding affinity of these proteins. Previous trials are currently under-utilized due to a lack of learning from these trials. Here, a new method is proposed which learns from previous trials to improve accuracy and reduce time and trials to obtain the candidate peptides and correct binding sites. This makes possible a better utilization of non-competitive inhibition. A framework for compiling the results of trials, learn from them, and utilize those learnings to predict the results of further docking trials is provided in the form of a graph database. Thus this new method harnesses the power of computation, pools knowledge from various trials and machine learning to aid the drug-discovery to combat diseases. To justify this algorithm, the docking tool Rosetta was used to find all possible orientations of 10 complexes taken from the RCSB PDB. Random orientations of Rosetta were sorted into T+, F-, F+ and T-, z-score threshold was chosen by plotting ROC curve of these 10 complexes. The CLICK tool was used to compare the input to the complexes in the database. It is shown that when the queried structure exists in the prediction database, an accurate prediction can be expected. This justifies the algorithm used in this experiment. When the queried structure does not exist in the database, the prediction cannot be entirely relied upon. The predicted structure in this case is shown superimposed with the crystal structure in the superimposed image. The structure of the MglA-MglB complex was predicted and verified in vitro.