INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

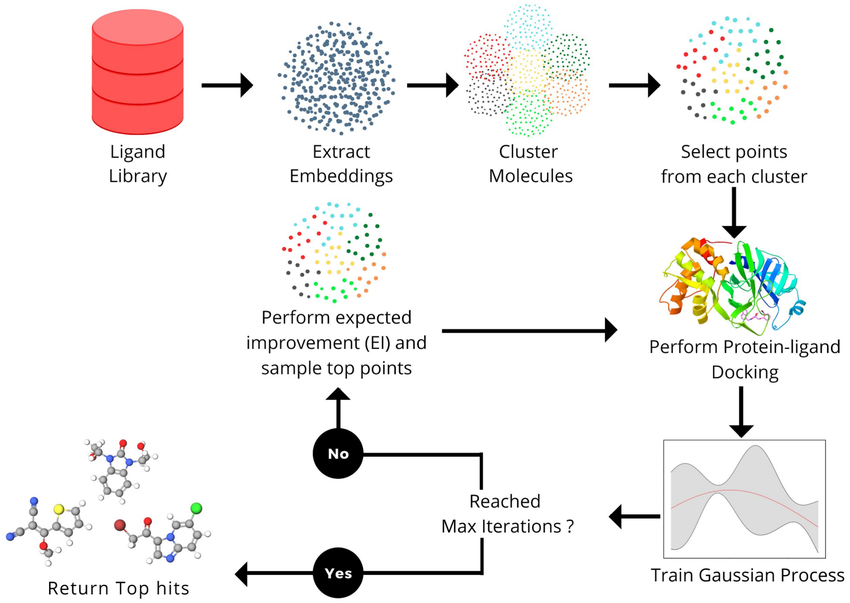
R&D SHOWCASE 2021
Technology, Social Impact

Enhanced Sampling of Chemical Space for High Throughput Screening Applications using Machine Learning

ABSTRACT

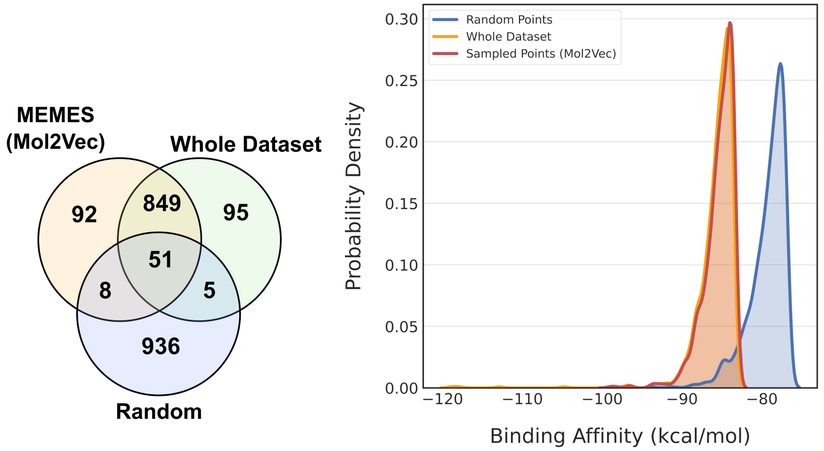
In drug discovery applications, high throughput virtual screening exercises are routinely performed to determine an initial set of candidate molecules referred to as “hits”. In such an experiment, each molecule from large small-molecule drug library is evaluated for physical property such as the binding affinity (docking score) against a target receptor. In real-life drug discovery experiments, the drug libraries are extremely large but still a minor representation of the essentially infinite chemical space , and evaluation of physical property for each molecule in the library is not computationally feasible. A novel machine learning framework “MEMES” based on Bayesian optimization is proposed for efficient sampling of chemical space. The proposed framework is demonstrated to identify 90% of top-1000 molecules from a molecular library of size about 100 million, while calculating the docking score only for about 6% of the complete library.

METHOD



Overview of the proposed MEMES framework

RESULTS



(a) (b)

Figure shows the performance of MEMES framework on Ultra Large Docking Library (~100 million compounds) against the AmpC target receptor.

Authors: Sarvesh Mehta, Siddhartha Laghavarapu, Yashaswi Pathak, Aaftaab Sethi, Mallika Alvala, U Deva Priyakumar

Research Center Name:----------