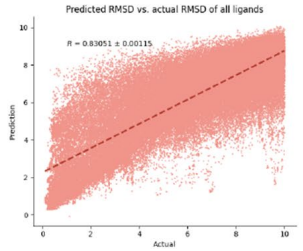
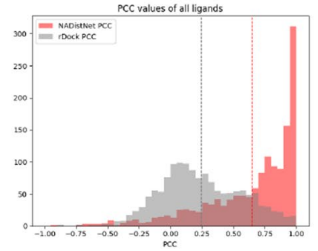
RmsdXNA: A platform for RNA/DNA-targeting Virtual Screening using Machine-Learning Method

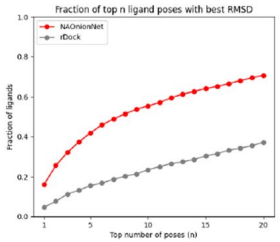
- **Model: For regression prediction, the XGBoost algorithm is chosen due to its fast training speed and strong predictive capabilities.**
- **Feasures: NA-ligand distances in the physical meaningful forms: 1/r and 1/r^6**



(a) Scatterplot of combined actual RMSD vs. NADistNet's predicted RMSD of validating dataset in each fold. Correlation coefficient (R) of 0.82950 ± 0.00115 is obtained.



(b) PCC distribution of ligand's poses RMSD vs. rDock score (grey) and NADistNet RMSD prediction (red). NADistNet has an average PCC of 0.64531 ± 0.36262, whereas the rDock has a PCC of 0.24361 ± 0.32229.



(c) Fraction of ligands where the top n poses selected by the scoring method contains the pose with lowest RMSD.

- **Target: RNA target**
- **Library: Chembridge Corp screening libraries (1.5M)**
- **Validated by MD simulations**

OXFORD

Briefings in Bioinformatics, 2024, 25(3), bbae166
https://doi.org/10.1093/bib/bbae166
Problem Solving Protocol

RmsdXNA: RMSD prediction of nucleic acid-ligand docking poses using machine-learning method

Lai Heng Tan, Chee Keong Kwoh and Yuguang Mu