Prioritizing virtual screening with interpretable interaction fingerprints

Alexandre Victor Fassio [1] Laura Shub [2] Luca Ponzoni [2] Jessica Mckinley [2] Matthew O'Meara [3] Rafaela Salgado Ferreira [4] Michael Keiser [2] Raquel Minardi [5]

Vol 1, 2022 - 152537
Poster

Abstract

Machine learning-based (ML) drug discovery success depends on molecular representation. Yet traditional molecular fingerprints omit the protein and pointers back to structural information that would enable better model interpretability. Therefore, we propose LUNA, a Python 3 toolkit that calculates and encodes protein-ligand interactions into new hashed fingerprints inspired by Extended Connectivity FingerPrint (ECFP): EIFP (Extended Interaction FingerPrint), FIFP (Functional Interaction FingerPrint), and Hybrid Interaction FingerPrint (HIFP). LUNA also provides visual strategies to make the fingerprints interpretable. We performed three major experiments exploring the fingerprints' use. First, we trained ML models to reproduce DOCK3.7 scores using 1 million docked Dopamine D4 complexes. We found that EIFP-4,096 performed (R^2 = 0.61) superior to related molecular and interaction fingerprints. Secondly, we used LUNA to support interpretable ML models. Finally, we demonstrate that interaction fingerprints can accurately identify similarities across molecular complexes that other fingerprints overlook.

Institutions

[1] USP - São Carlos

[2] University of California San Francisco

[3] University of Michigan

[4] UFMG (Federal Universtity of Minas Gerais)

[5] Universidade Federal de Minas Gerais

Keywords

interaction fingerprints

protein-ligand interaction patterns

interpretable fingerprints

compound prioritization

interaction visual analysis