Cluster-Guided Large-Scale Virtual Screening of Natural Products for Antiviral Drug Discovery

Name MUHAMMAD ALQAAF SUBANDOKO

Laboratory' s Computational Systems Biology

Supervisor' s name Shigehiko KANAYA

Abstract (should be within 1st page)

Viral infectious diseases, including SARS-CoV-2 and HIV-1, pose significant global health challenges. Although natural products provide structurally diverse antiviral scaffolds, navigating extensive metabolite databases is computationally demanding, and structure-based screening is often limited by target redundancy. This dissertation introduces a cluster-guided virtual screening framework designed to systematically prioritize natural product-derived antiviral candidates. Secondary metabolites from the KNApSAcK database were mapped to pertinent viral targets using BindingDB. Structural redundancy was addressed through sequence similarity clustering, consolidating related viral variants into representative groups. The framework incorporates high-throughput molecular docking alongside stringent pharmacokinetic filtering to ensure the optimal drug-like properties of the identified compounds. Candidates were subsequently benchmarked against clinical drugs using comparative statistical methods, such as the Mann–Whitney U test, Welch's t-test, and effect size quantification, to mathematically validate their binding affinities. When applied to SARS-CoV-2 and HIV-1, the methodology effectively reduced redundancy and identified multiple natural product metabolites with predicted affinities that were competitive with or superior to those of established drugs.

Keywords: Antiviral drug discovery; Cluster-guided virtual screening; Natural products; Sequence-similarity clustering; Molecular docking; ADME profiling; SARS-CoV-2; HIV-1