PURDUE INNOVATES | Office of Technology Commercialization

Software to generate mega chemical library

Non-toxic peptides that selectively degrade misfolded proteins for neurodegenerative diseases.

Early drug discovery is a very time-consuming and costly process. To accelerate drug discovery, scientists often utilize computational tools to search for compounds that have potential to become pharmaceutically active molecules . Nowadays, public chemical databases or libraries typically contain 10 billion or few compounds. However, researchers at Purdue University have developed software to generate mega chemical libraries. Unlike other software products, this solution is not based on artificial intelligence or machine learning but is rather a systems and knowledge driven approach to identify potential lead molecules. The approach starts from a given compound and readily generates billions or trillions of compounds, covering an unprecedentedly large chemical space, which may contain improved candidates for drug discovery. More specifically, this fragment module library invented by Purdue scientists uses a fully automated approach that is composed of the following steps - 1) creation of backbone and peripheral libraries, 2) introduction of chemical mutations, 3) systematic combination of mutated backbone constituents and peripheral constituents, and 4) systematic attachment of formulated modules to the parent compound - to create a mega library of potential lead compounds that have high potentiality for biological activity and high synthesis accessibility.

Technology Validation:

The fragment module library generated using an example software program of the disclosure (viz., ChemHopperTM) versus Deepfrag was able to generate more known lead candidates when tested on three known small molecules.

Advantages:

-Fully automated

Technology ID

2025-LI-70824

Category

Biotechnology & Life Sciences/Biomarker Discovery & Diagnostics

Biotechnology & Life Sciences/Analytical & Diagnostic Instrumentation

View online page

-Starts from a given compound and readily generates billions or trillions of compounds

-Extremely large library size compared to current public libraries

Applications:

-Synthetic biology

-Pharmaceuticals

-Drug Discovery

TRL: Biotechnology

Intellectual Property:

Provisional-Patent, 2024-09-13, United States

Utility Patent, 2025-09-12, United States

Keywords: Computer Technology, Drug Discovery, Fragment libraries, Lead optimization, Pharmaceuticals