[SS3-O-269]

Development of a Machine Learning Model for the Synthesis and Molecular Docking Study of HIV Integrase Inhibitors

Tuyet-Minh Phan[1], Thao-Linh Song Nguyen[1], Bang-Vu Duong[1], Thanh-An Pham[1], The-Chuong Trinh[1], Van-Thinh To[1], Tieu-Long Phan[1], and Tuyen Ngoc Truong[1]*

[1]*Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, 41-43 Dinh Tien Hoang Street, District 1,*

Ho Chi Minh city, Vietnam.

*Corresponding author: Tuyen Ngoc Truong, Assoc. Prof. PhD

Presenter: Tuyet-Minh Phan

Email: ptminh.ths.dldls23@ump.edu.vn

Purpose This study aims to develop machine learning and molecular docking models for the screening of in-house databases and the synthesis of potential HIV integrase inhibitors.

Methods This study extracted 2126 compounds from the ChEMBL 33 database and calculated their molecular descriptors using the Mordred library. The data was then subjected to a rigorous machine learning pipeline, including missing values and outliers handling, feature selection, and model selection, to develop a robust QSAR model. This model was subsequently used to predict the in-house library, which consists of synthesized triaryl scaffolds, and was validated using molecular docking with Autodock-GPU. The candidate compounds were then synthesized through a four-step process: (i) Williamson reaction, (ii) hydrazone formation, (iii) Vilsmeier-Haack reaction, and (iv) reductive amination reaction.

Results The Extra Tree algorithm was the optimal candidate during feature selection, achieving an F1 score of 0.783±0.050 in cross-validation. The MLP and TomekLinks algorithms were selected to develop the predictive model, with a cross-validated F1 score of 0.824±0.043 for the model development and optimization step. The external validation also proved the model's generalizability, with the F1 score being 0.845. The screening phase proposed compound DI081 as a potential candidate, with a 98.57% inhibitory probability and a -20.77 kcal/mol docking score from Autodock-GPU, surpassing Bictegravir's -10.58 kcal/mol. Finally, DI081 was synthesized with a 2.43% yield of the entire procedure due to steric hindrance from 2-hydroxy acetophenone. The proposed chemical structure is anticipated to exhibit potential inhibitory activity against HIV integrase. This hypothesis is currently being evaluated through rigorous *in vitro* testing to prove its efficacy.

Conclusions This study has identified DI081 as a compound with potent HIV integrase inhibitory activity, as evidenced by its predicted probability of 98.57% and docking score of -20.77 kcal/mol

KEYWORDS: HIV integrase, machine learning, molecular docking, synthesis, 1*H*-pyrazole.

Graphical abstract