DESIGN OF NOVEL MOLECULES WITH ANTI-COVID-19 ACTIVITY: FROM COMPUTER SIMULATIONS TO ORGANIC SYNTHESIS

Lohachova K.O., Yevsieieva L., Zakharov A.B., Ivanov V.V.,
Kyrychenko A.V., Kovalenko S.M., Kalugin O.N.
V. N. Karazin Kharkiv National University, sq. Svobody, 4, Kharkiv, 61022, Ukraine
e-mail: **vivanov@karazin.ua**

I. COMPUTER-AIDED DRUG DESIGN OF NOVEL NIRMATRELVIR ANALOGS INHIBITING MAIN PROTEASE OF CORONAVIRUS SARS-COV-2

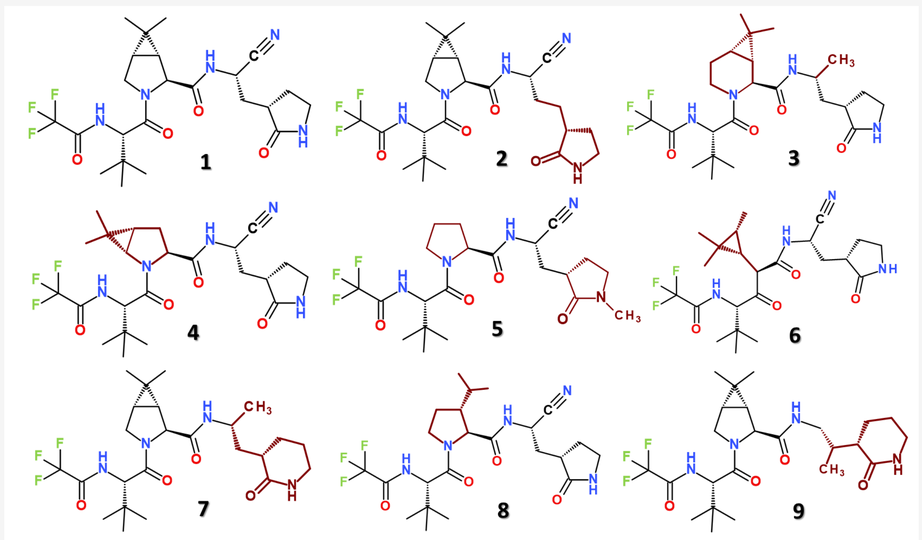


Fig. 1. Nirmatrelvir molecule (1) and its designed analogs

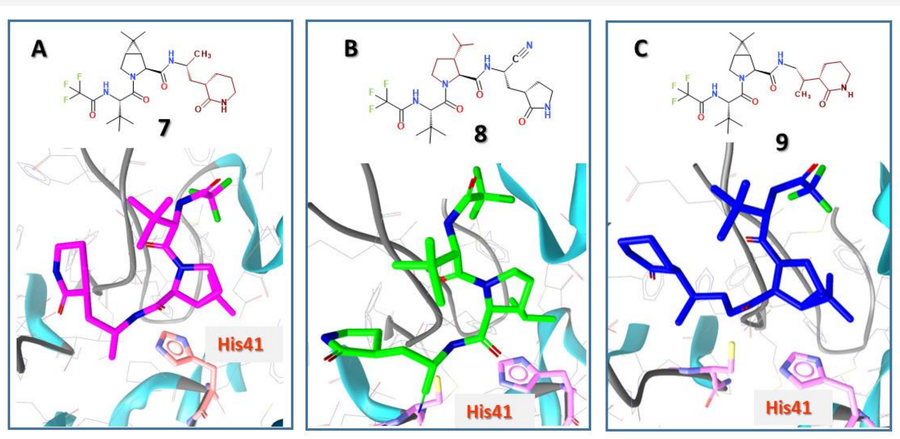


Fig. 2. Binding modes of some nirmatrelvir's analogs at the active site of the M^{pro} enzyme, estimated by molecular docking calculations

II. DATA-DRIVEN DISCOVERY OF FUNCTIONAL MATERIALS WITH ANTI-COVID-19 AND OTHER ACTIVITIES:LARS–LASSO LOGISTIC REGRESSION FOR QSAR/QSPR DESIGN OF COMPOUNDS

The possibility of using the L1-regularization to obtain logistic classification equations of quantitative/qualitative structure-activity/property relationships (QSAR/QSPR) have been investigated. The least angle regression (LARS) of least absolute shrinkage and selection operator (LASSO) variant has been implemented in the logistic regression. The method was used for building simple classification functions for prediction activity against COVID-19 main protease.

LARS-LASSO logistic regression:

$$p_i(\beta) = 1/(1+\exp(-f_i)), \qquad f_i = \beta_0 + \sum_j \beta_j x_{ij}, \qquad s.t. \quad \|\beta\|_1 < \xi$$

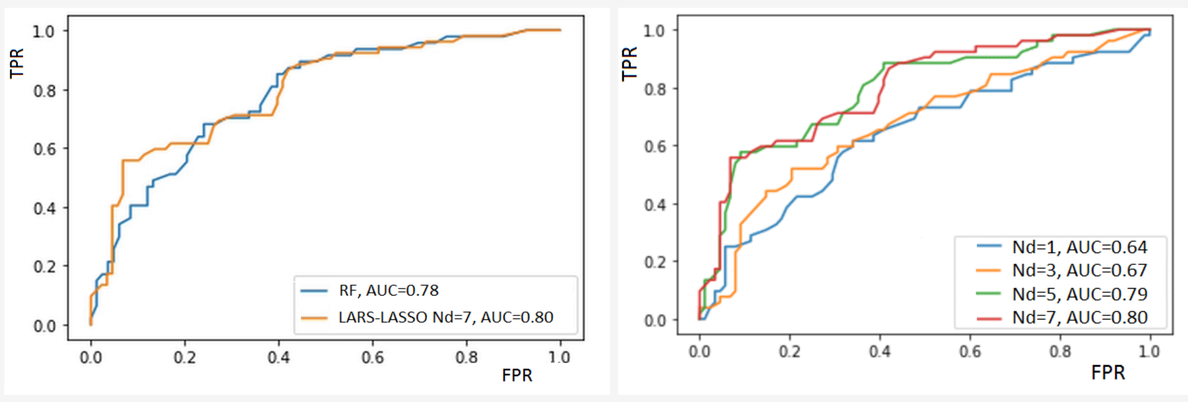


Fig. 3. The ROC-curve (true positive rate, TPR *vs* false positive rate FPR) obtained for anti-COVID test-set molecules classification.

III. FROM ML300 TO NOVEL NON-COVALENT INHIBITORS OF SARS-COV-2 MAIN PROTEASE *VIA* EVOLUTIONARY *DE NOVO* DESIGN, VIRTUAL SCREENING, MOLECULAR DYNAMICS, AND RETROSYNTHETIC STRATEGIES

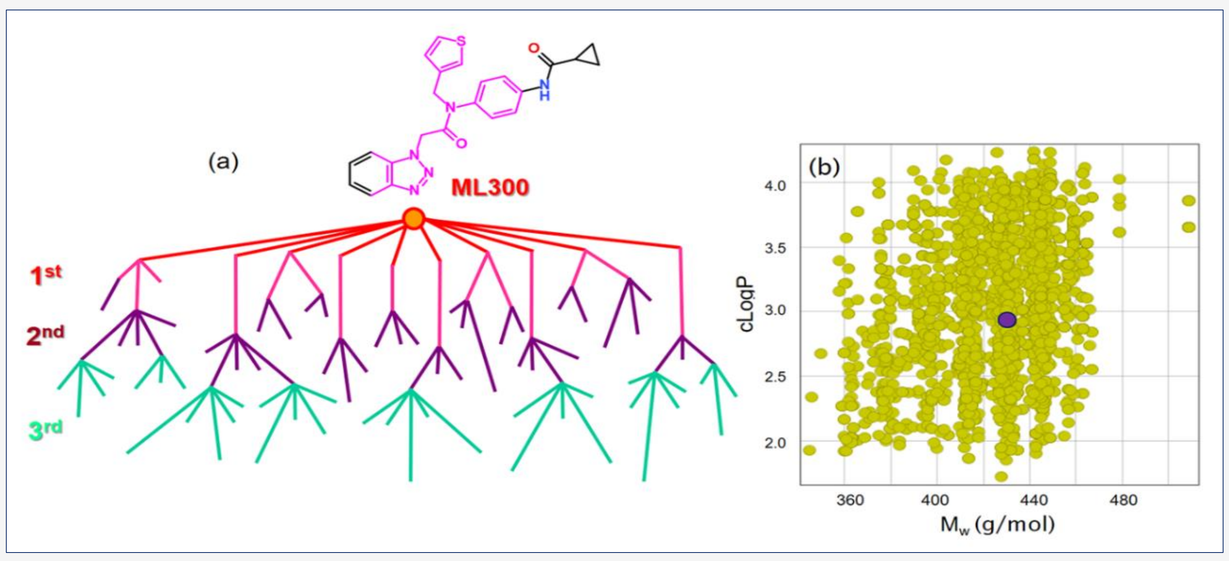


Fig. 4. Scheme of generation of an evolution library of **ML300** analogues

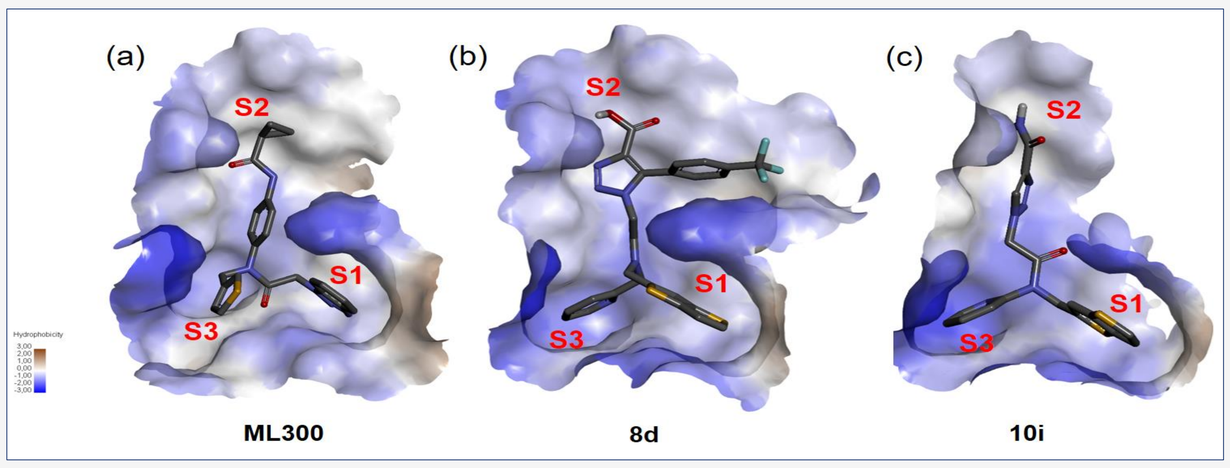
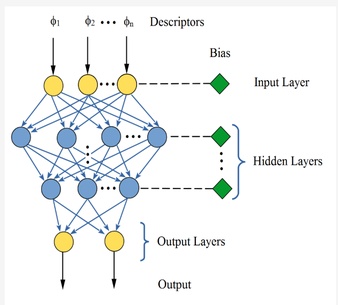


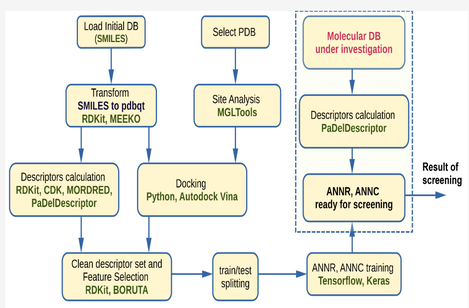
Fig. 5. Hydrophobic surface of the ligand-M^{pro} complexes, as estimated by molecular docking against M^{pro} (PDB: 7LME)

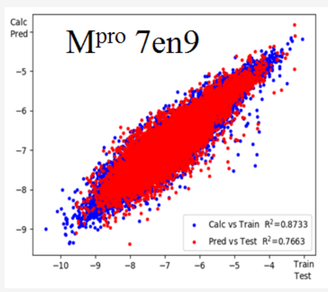
IV. NEURAL NETWORK BASED MODELING AND SCREENING OF CANDIDATE M^{PRO} AND PL^{PRO} INHIBITORS

Steps of NN tuning:

1) Virtual library preparation (~64000 molecules); **2**) Download crystal structure PDB files: M^{pro} (7en9) Pl^{pro} (7lbr); **3**) Formation of a binding site (**MGLTools**); **4**) Fast docking (**Python, Autodock Vina**); **5**) Molecular Descriptors calculations (**Python, RDKit, PaDELDescriptors**); **6**) Clean the set of descriptors and feature selection (**Rdkit, BORUTA**); **7**) Splitting the library into ***train*** (70%) and ***test*** (30%) (**Python, Scikit-learn**); **8**) NN training (**Tensorflow, Keras**);







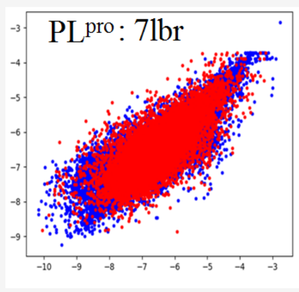


Fig. 6. Evaluation the accuracy of regression and classification. The NN training for the Mpro model demonstrated a coefficient of determination R^2(train)=0.87, R^2(test)=0.78.

For the NN classification model (active systems correspond to affinity more then 7 kcal/mol. Area under ROC curves: AUC_{train}=0.95 and AUC_{test}=0.90. These results demonstrate strong predictive performance.