CONSORTIUM USTH VIETNAM FRANCE UNIVERSITY

ICCS

ĐẠI HỌC Y DƯỢC THÀNH PHỐ HỒ CHÍ MINH VIỆT NAM 1947

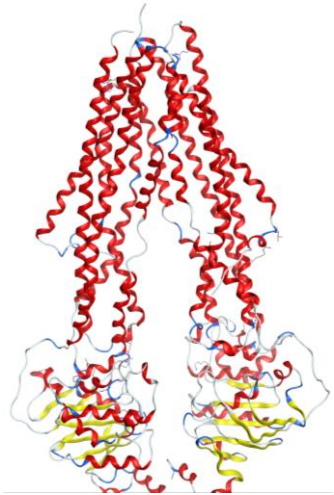
KHOA DƯỢC Đại học Y Dược Thành phố Hồ Chí Minh

1st International Conference on Chemical Science

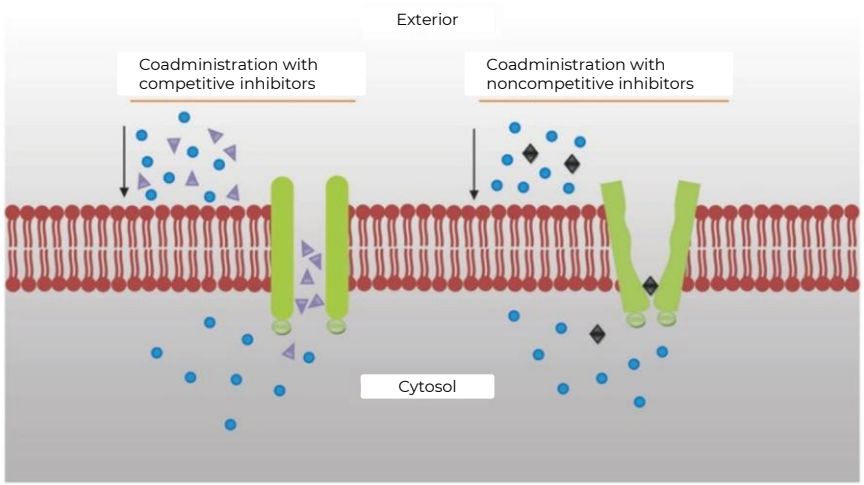
DEEP-LEARNING CLASSIFICATION MODEL ON SCREENING P-GLYCOPROTEIN INHIBITORS

Presenting author: **Vo Thanh Phuong**

Multidrug resistance and P-glycoprotein



P-glycoprotein
(PDB code: 7A69)



P-glycoprotein inhibitors

G AS, Jodie Y, Andrew W, et al. *Science (1979)*. 2009;323(5922):1718-1722
Amin MdL. *Drug Target Insights*. 2013;7(1).

P-glycoprotein inhibitors

1st generation

- Substances with clinical pharmacological effects
- High serum concentration
- High toxicity
- Nonspecific

2nd generation

- Structural transformation from 1st generation.
- Neutralizes the original pharmacological effect
- Reduce toxicity
- Nonspecific

3rd generation

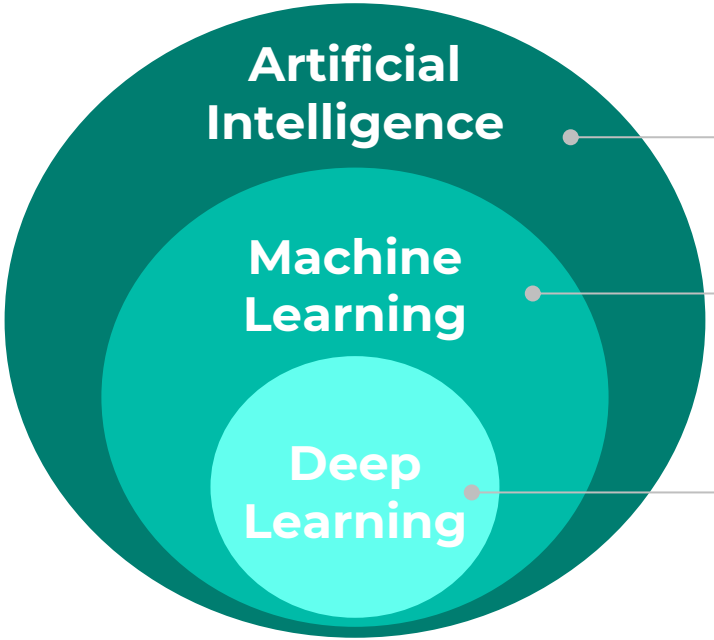
- Low serum concentration
- Low toxicity
- Higher specific to P-gp

Ex: elacridar, zosuquidar, taxquidar

4th generation

- Natural compounds
- Low toxicity
- More efficient.
- Higher specific to target.

Artificial Intelligence (AI)



- Simulate human intelligence

- Relatively small amount of data
- Important attribute → Human

- Huge amount of data
- Important properties → Model classes, types and complexity

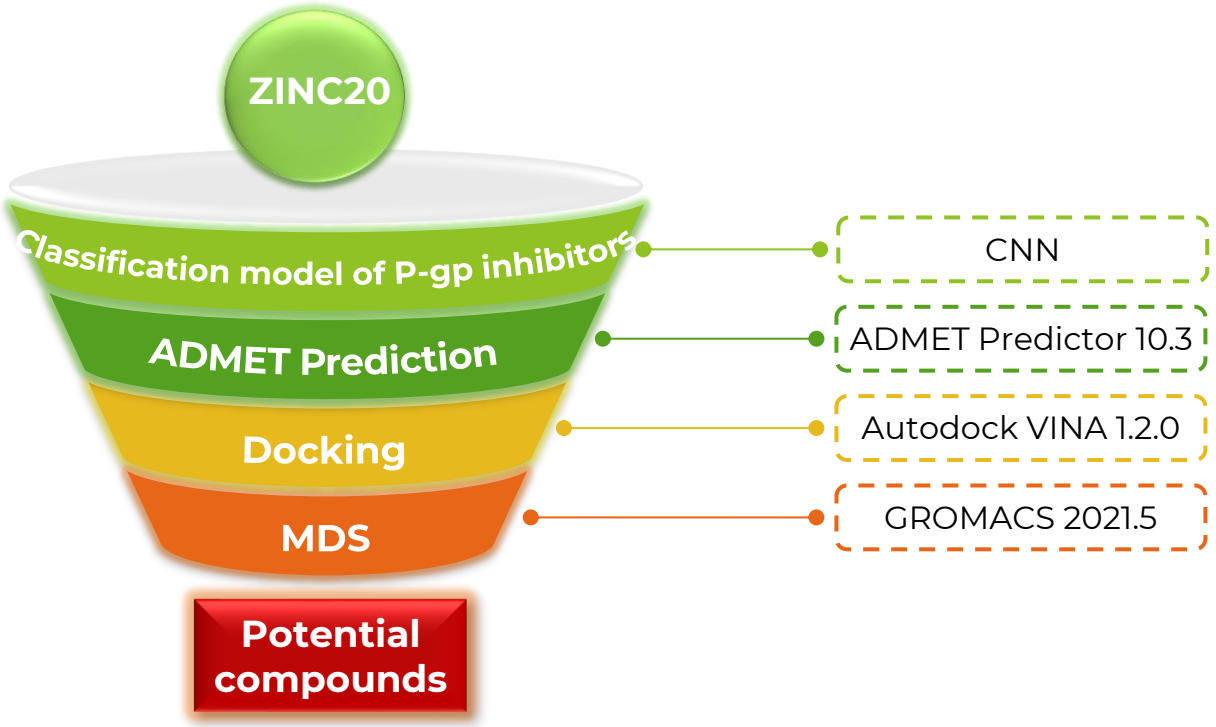
The study questions

Inhibitors

Non-inhibitors

P-GLYCOPROTEIN

Methodology Flowchart



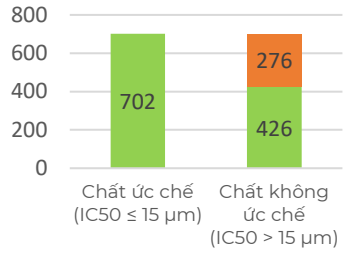
Classification models were published

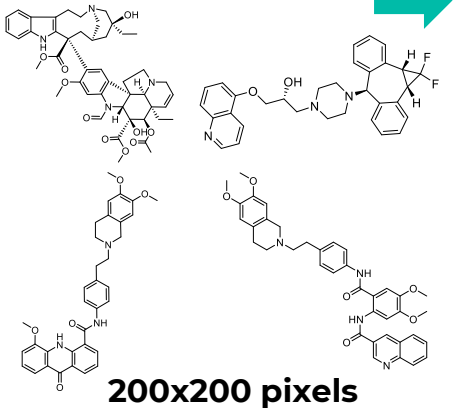
Research	Year	Methods	Databases	Accuracy
H.Sun et al	2005	NB	609 compounds	82.2%
YH.Wang et al	2005	Kohonen	206 compounds	82.3%
		BPNN	174 compounds	29%
P.Crivori et al	2006	PLSD	148 compounds	82%
F.Broccatelli et al	2011	MIF	1195 compounds	86%
L.Chen et al	2011	DT, RF, NB	1273 compounds	81.2%
S.Rapposelli et al	2012	RF, C4.5, consensus	59 compounds	75% (RF)
V.Poongavanam et al	2012	RF, kNN, SVM	1935 compounds	75% (RF)
F.Klepsch et al	2014	kNN, SVM, RF, DT, Binary QSAR	1954 compounds	73% (RF)
S.Eríc et al	2014	ANN, SVM, combine	135 compounds	89%
Khac-Minh Thai et al	2015	Backpropagation neural network	133 compounds	82%
V.Prachayasittikul et al	2015	DT, SVM, ANN	2477 compounds	87,4% (ANN)
M.Yang et al	2015	18 models machine learning	2428 compounds	77.4-79.9%
Khac-Minh Thai et al	2016	BPNN, C5.0, C&R Tree, QUEST, CHAID, Logistic Regression, Bayesian Network, Discriminant Analysis, SVM, combine	2131 compounds	92% (validation set) 100% (test set)
VK.Hinge et al	2019	GBM, GLM, SVM, kNN	1274 compounds	87.1-96.9% (SVM)
This research	2022	CNN	1404 compounds	-

Classification model of P-gp inhibitors

Databases

1404 compounds

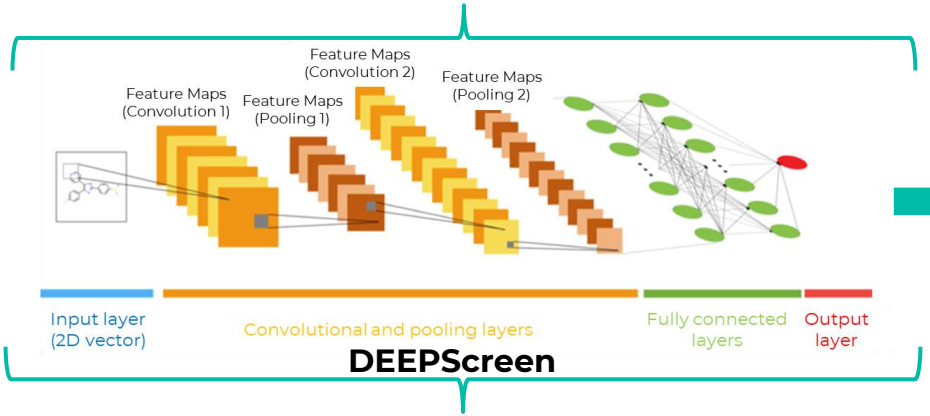




200x200 pixels

Optimization algorithm and hyperparameters

Optimization algorithm	Learning rate	Epoch	Fully connected layer	Drop out rate



DEEPScreen

- Inhibitor
- Non-inhibitor

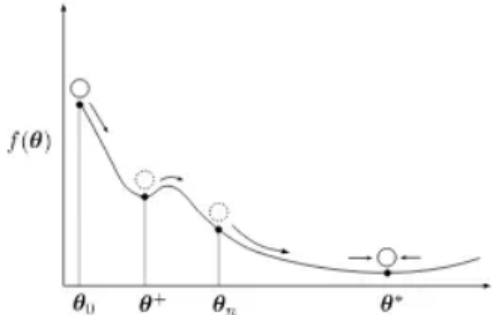
Evaluated parameters

Precision	Accuracy	Recall	F1score	MCC	AUC

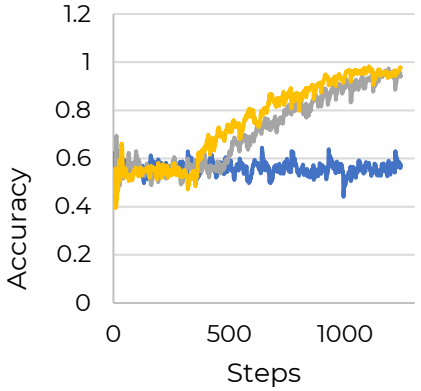
Evaluating optimization algorithm and hyper-parameters

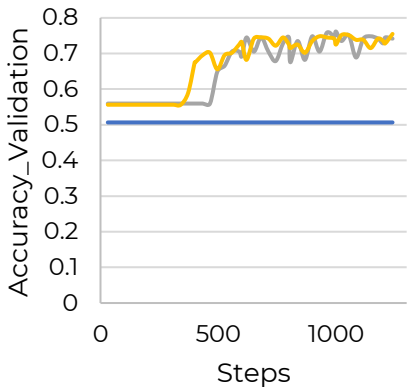
Parameters	Optimization algorithm	Learning rate	Epoch	Fully connected layer	Drop out rate
Optimization algorithm	-	0.0005	40	16	0.6
Learning rate	adam	-	40	16	0.6
Epoch	adam	0.0005	-	16	0.6
Fully connected layer	adam	0.0005	40	-	0.6
Drop out rate	adam	0.0005	40	16	-

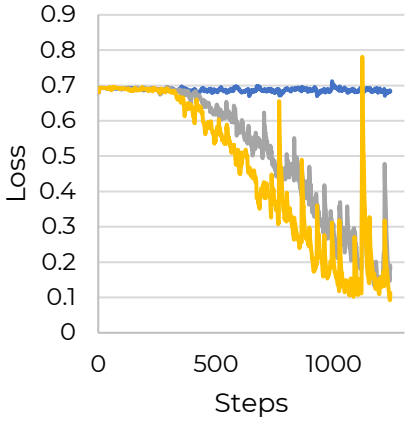
Optimization algorithm

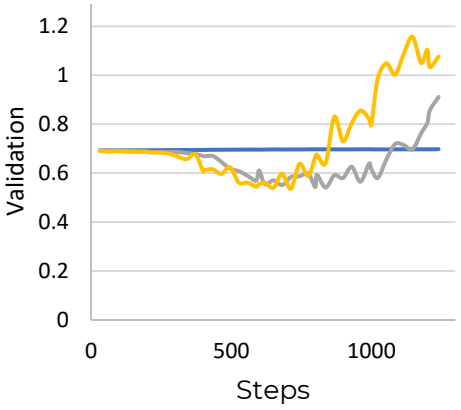


Parameters	Precision	Accuracy	Recall	F1score	MCC	AUC
Mometum	0.52	0.0	0.0	0.0	0.0	0.52
	0.49	0.0	0.0	0.0	0.0	0.53
	0.42	0.0	0.0	0.0	0.0	0.49
RMSprop	0.72	0.71	0.85	0.77	0.43	0.83
	0.74	0.73	0.82	0.77	0.47	0.80
	0.74	0.82	0.70	0.76	0.49	0.82
Adam	0.76	0.78	0.80	0.79	0.50	0.86
	0.77	0.75	0.88	0.81	0.53	0.85
	0.76	0.77	0.85	0.81	0.51	0.81



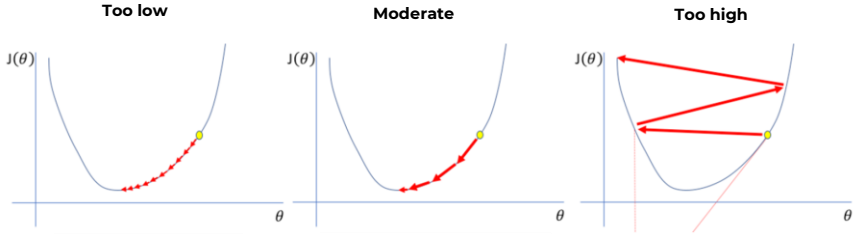




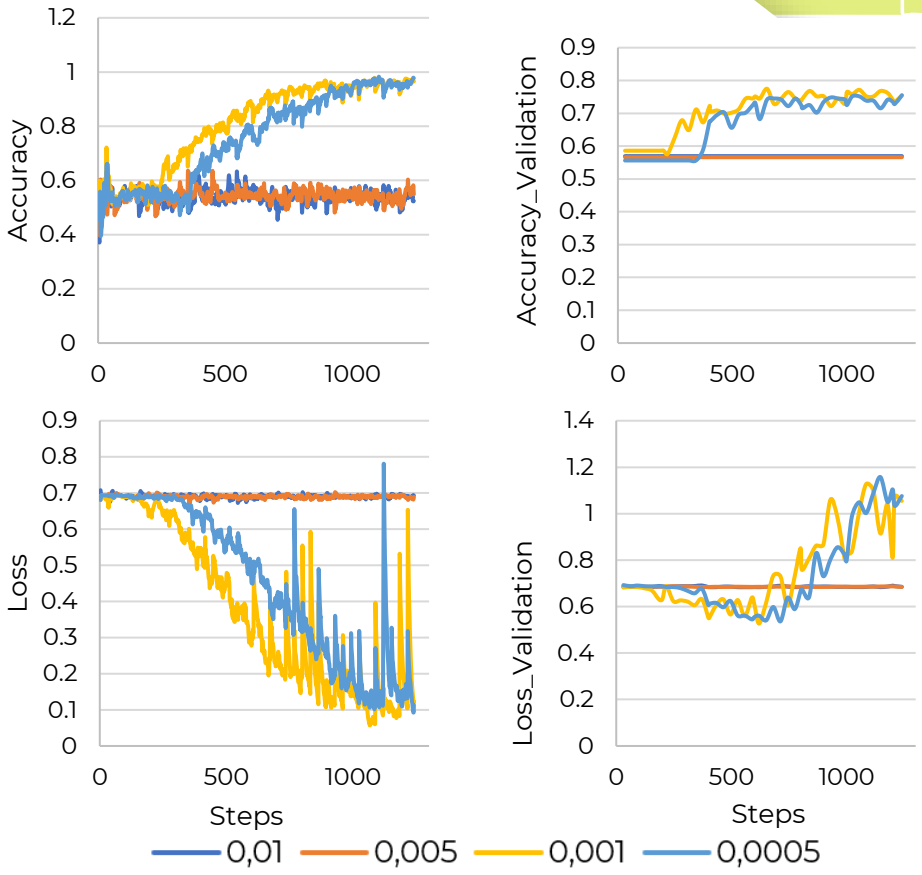


momentum RMSprop adam

Learning rate

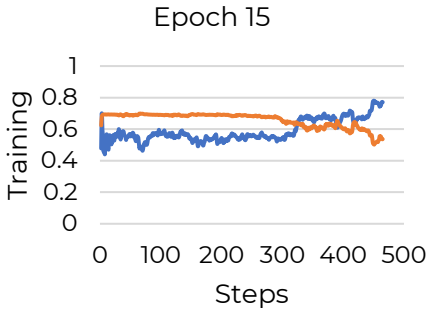


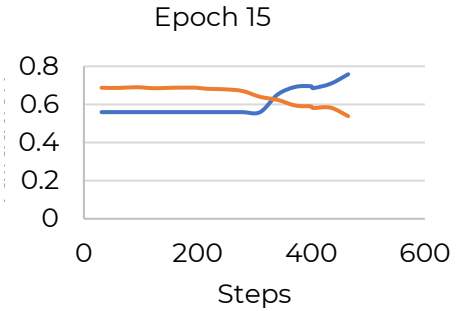
Learning rate	Precision	Accuracy	Recall	F1score	MCC	AUC
0.01	0.45	0.0	0.0	0.0	0.0	0.50
	0.46	0.0	0.0	0.0	0.0	0.50
	0.48	0.0	0.0	0.0	0.0	0.50
0.005	0.45	0.0	0.0	0.0	0.0	0.50
	0.42	0.0	0.0	0.0	0.0	0.50
	0.46	0.0	0.0	0.0	0.0	0.50
0.001	0.73	0.75	0.73	0.74	0.46	0.81
	0.74	0.71	0.84	0.77	0.47	0.81
	0.72	0.68	0.88	0.77	0.45	0.84
0.0005	0.76	0.78	0.80	0.79	0.50	0.86
	0.77	0.75	0.88	0.81	0.53	0.85
	0.76	0.77	0.85	0.81	0.51	0.81

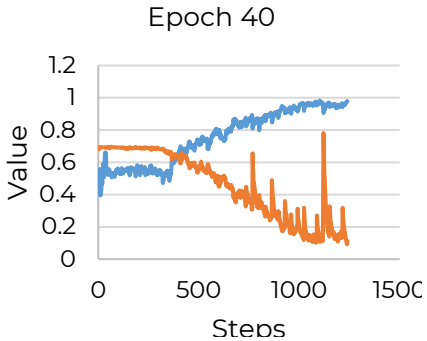


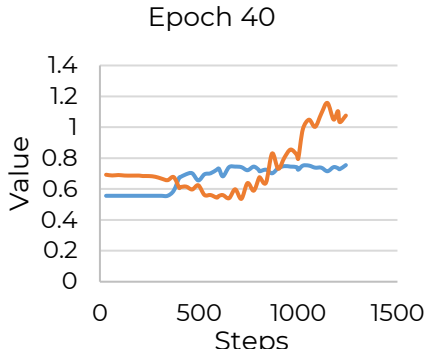
Epoch

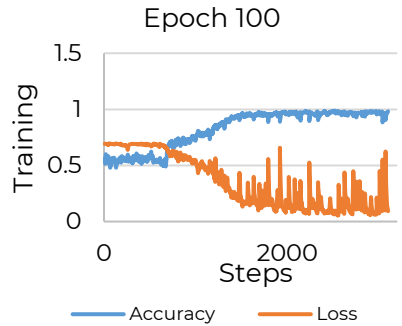
Epoch	Precision	Accuracy	Recall	F1score	MCC	AUC
5	0.48	1.00	0.02	0.04	0.09	0.73
	0.47	0.00	0.00	0.00	0.00	0.69
	0.45	0.00	0.00	0.00	0.00	0.72
15	0.69	0.69	0.77	0.73	0.37	0.77
	0.71	0.75	0.76	0.75	0.41	0.79
	0.63	0.58	0.95	0.72	0.35	0.78
30	0.73	0.73	0.87	0.79	0.41	0.79
	0.74	0.79	0.70	0.74	0.48	0.81
	0.74	0.77	0.78	0.77	0.47	0.84
40	0.76	0.78	0.80	0.79	0.50	0.86
	0.77	0.75	0.88	0.81	0.53	0.85
	0.76	0.77	0.85	0.81	0.51	0.81
100	0.73	0.74	0.82	0.78	0.44	0.88
	0.75	0.78	0.79	0.79	0.48	0.81
	0.73	0.73	0.79	0.76	0.46	0.83

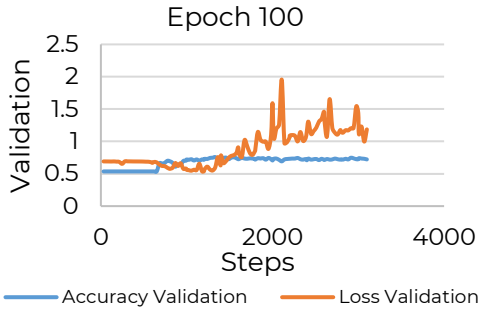






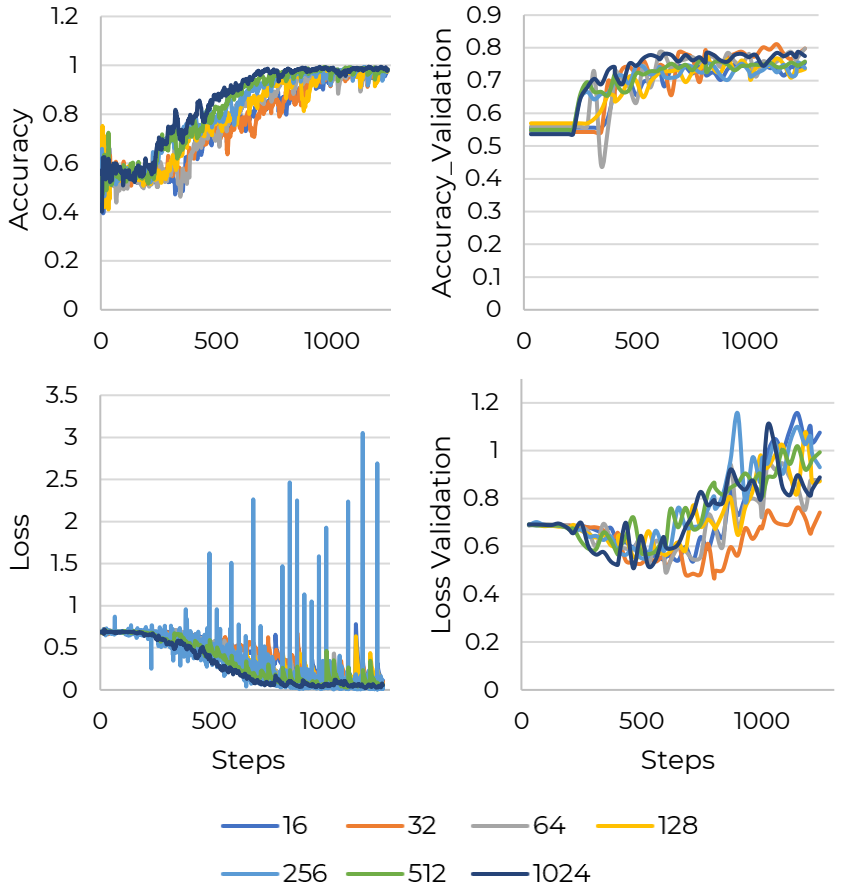




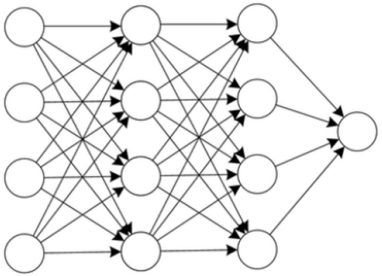


Fully connected layer

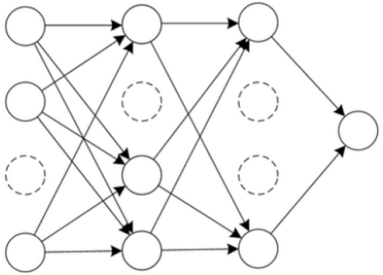
Fully connected layer	Precision	Accuracy	Recall	F1score	MCC	AUC
16	0.76	0.78	0.80	0.79	0.50	0.86
	0.77	0.75	0.88	0.81	0.53	0.85
	0.76	0.77	0.85	0.81	0.51	0.81
32	0.73	0.72	0.82	0.77	0.45	0.82
	0.78	0.77	0.83	0.8	0.56	0.84
	0.75	0.82	0.69	0.75	0.52	0.86
64	0.73	0.74	0.76	0.75	0.46	0.85
	0.75	0.74	0.88	0.8	0.46	0.85
	0.75	0.75	0.8	0.78	0.51	0.82
128	0.75	0.75	0.78	0.76	0.5	0.82
	0.73	0.73	0.76	0.75	0.45	0.81
	0.71	0.74	0.72	0.73	0.42	0.80
256	0.74	0.78	0.75	0.76	0.48	0.84
	0..73	0.78	0.75	0.76	0.45	0.84
	0.74	0.77	0.75	0.76	0.48	0.82
512	0.75	0.75	0.81	0.78	0.5	0.85
	0.73	0.68	0.91	0.78	0.46	0.82
	0.75	0.8	0.78	0.79	0.5	0.84



Drop out rate

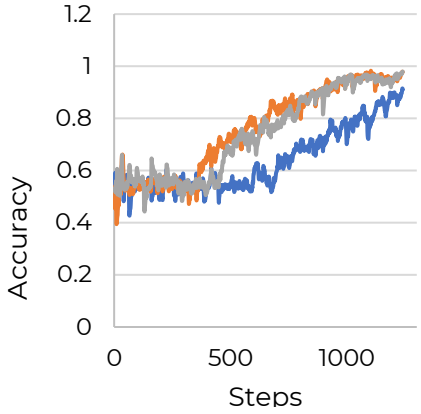


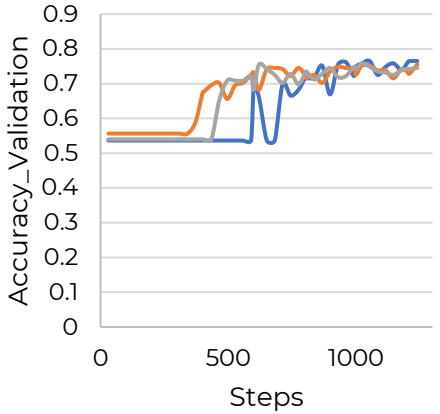
Normal neural network

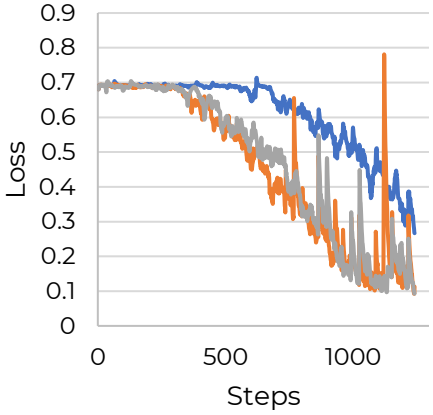


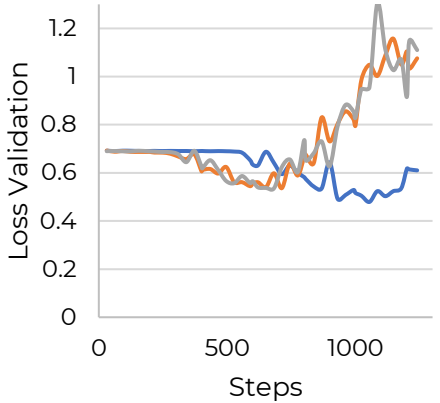
Neural network with drop out rate

Parameters	Precision	Accuracy	Recall	F1score	MCC	AUC
0.4	0.70	0.88	0.57	0.69	0.47	0.79
	0.74	0.69	0.86	0.77	0.49	0.83
	0.73	0.75	0.76	0.75	0.45	0.83
0.6	0.76	0.78	0.80	0.79	0.50	0.86
	0.77	0.75	0.88	0.81	0.53	0.85
	0.76	0.77	0.85	0.81	0.51	0.81
0.8	0.77	0.79	0.79	0.79	0.55	0.81
	0.69	0.71	0.71	0.71	0.38	0.83
	0.74	0.74	0.81	0.77	0.47	0.82





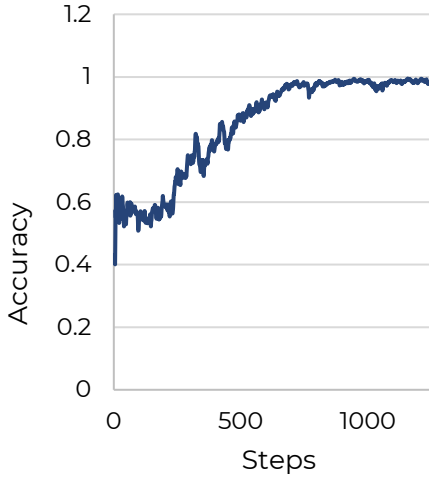


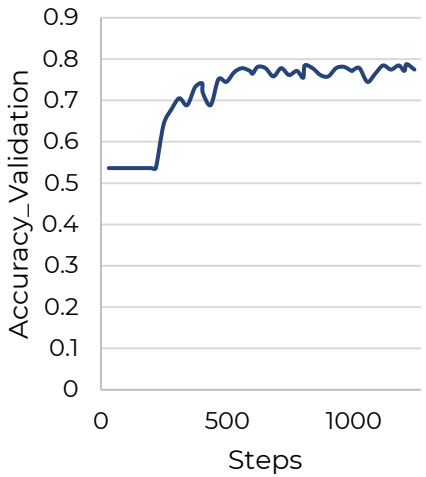


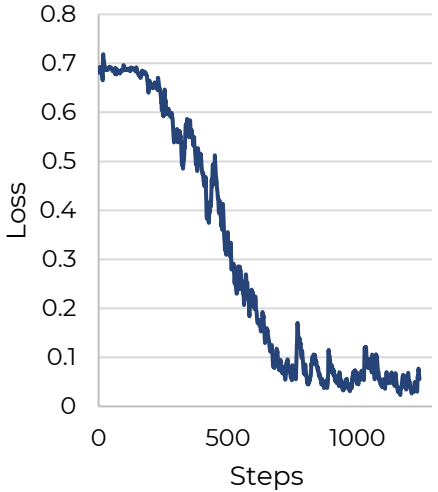
Selected model

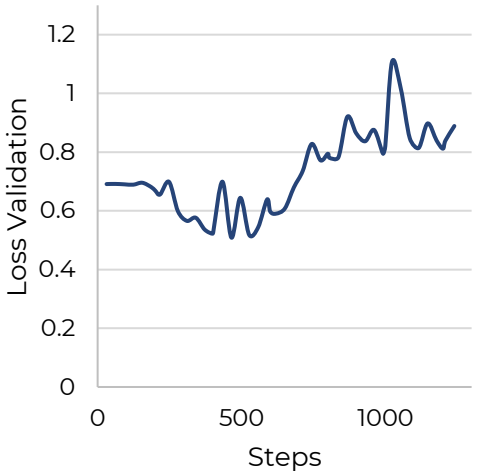
Optimization algorithm and hyperparameters | **Evaluated parameters**

Optimization algorithm	Learning rate	Epoch	Fully connected layer	Drop out rate	Precision	Accuracy	Recall	F1score	MCC	AUC
adam	0.0005	40	1024	0.6	0.79	0.75	0.91	0.82	0.58	0.86

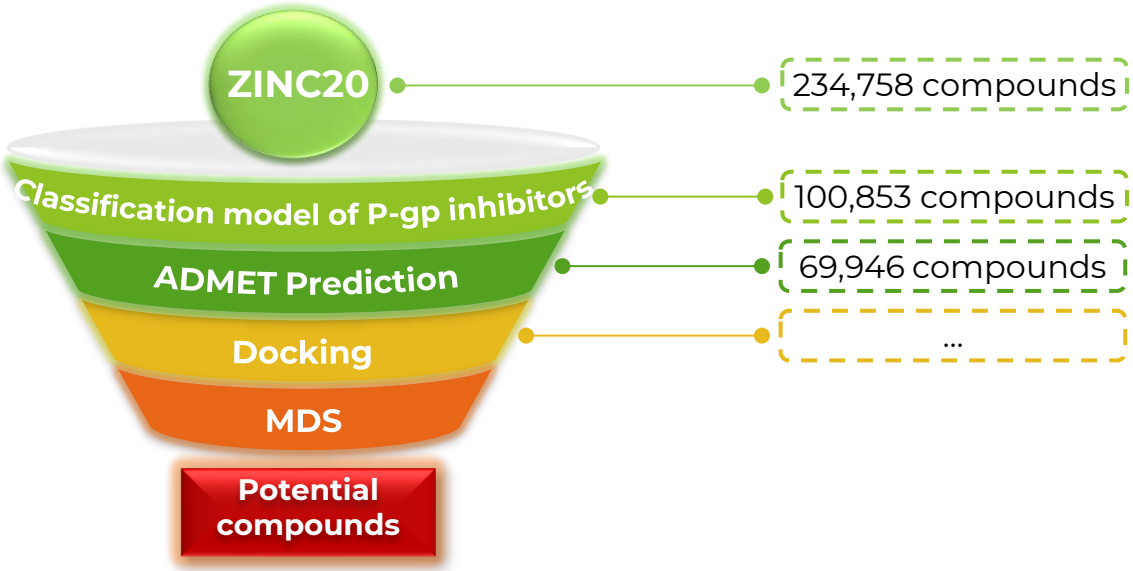




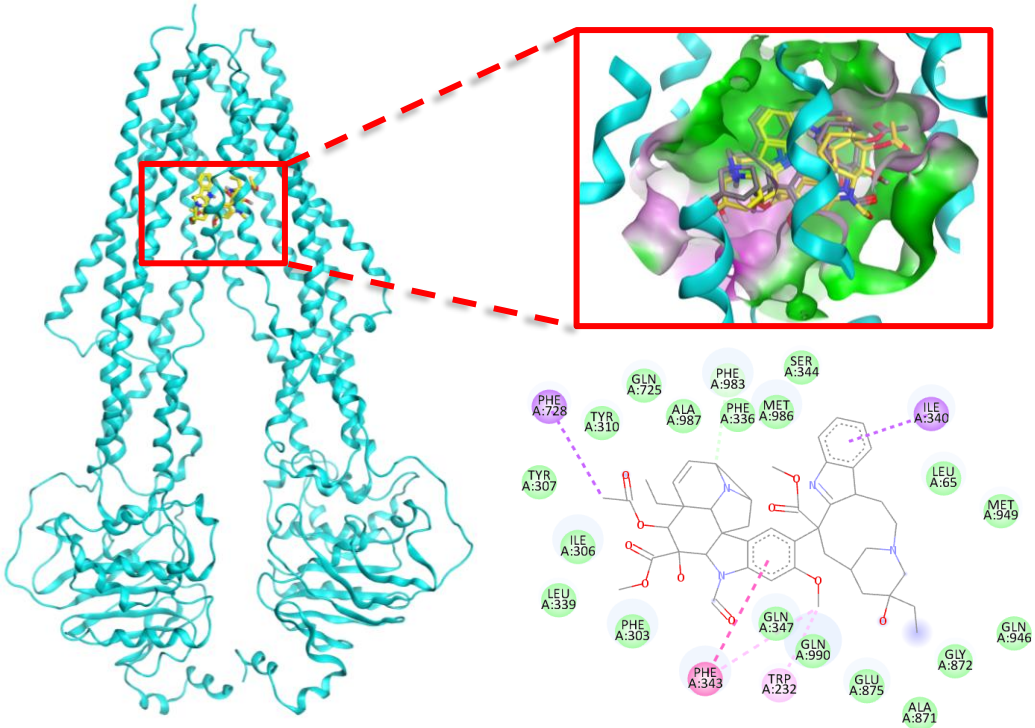




Result of screening P-gp inhibitors



Redocking

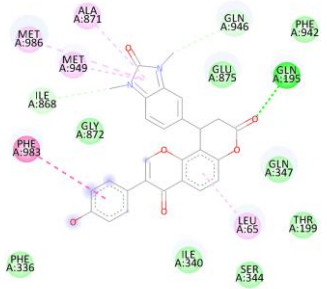


PDB code: 7A69
Resolution: 3.20 Å

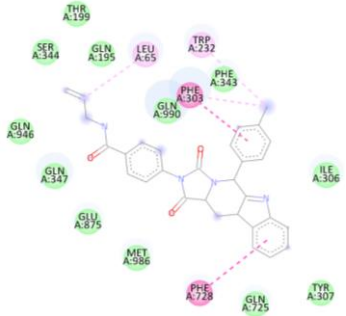
Vincristine

Docking score: -10.74 kcal/mol
RMSD value: 1.1334 Å

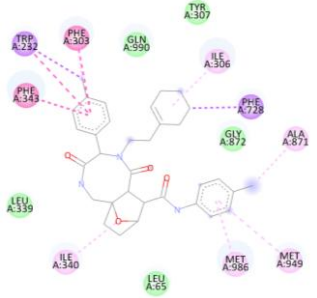
Docking



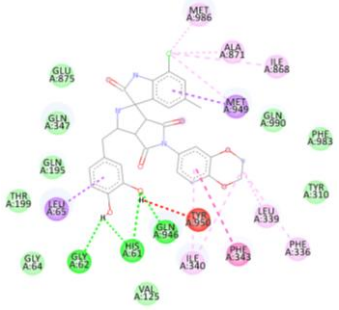
ZINC96221125 (A)
Docking score: -11.947 kcal/mol



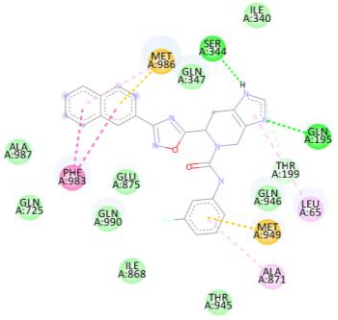
ZINC12882009 (B)
Docking score: -11.872 kcal/mol



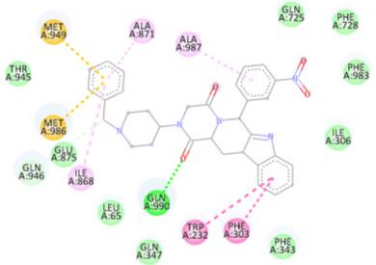
ZINC169771363 (C)
Docking score: -11.461 kcal/mol



ZINC70705366 (D)
Docking score: -11.313 kcal/mol



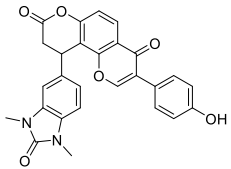
ZINC15675941 (E)
Docking score: -11.310 kcal/mol



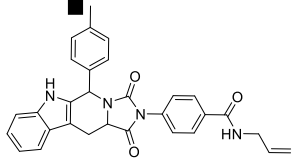
ZINC1877111 (F)
Docking score: -11.307 kcal/mol

RMSD – Root Mean Square Deviation

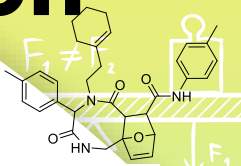
ZINC96221125 (A)
Docking score: -11.947 kcal/mol



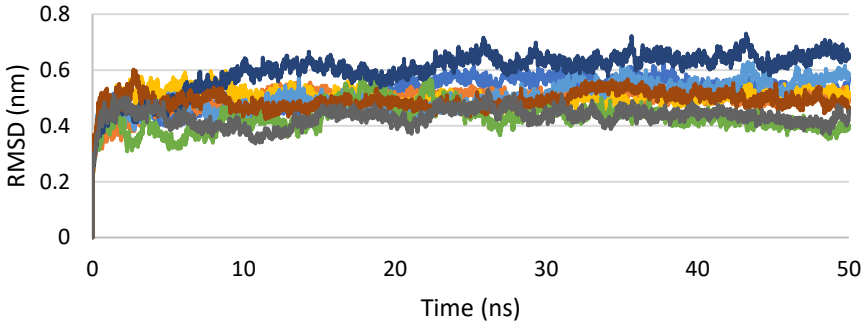
ZINC12882009 (B)
Docking score: -11.872 kcal/mol



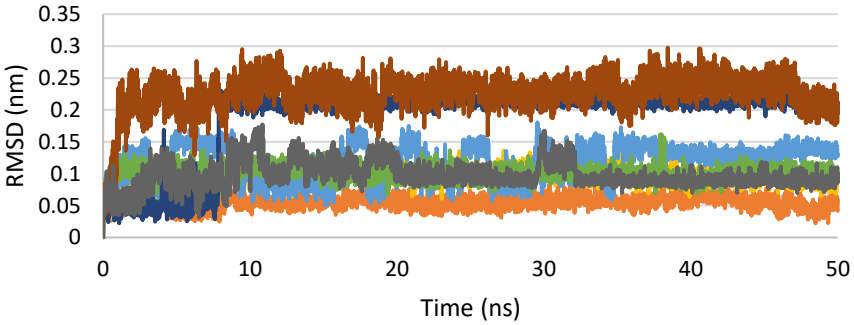
ZINC169771363 (C)
Docking score: -11.461 kcal/mol

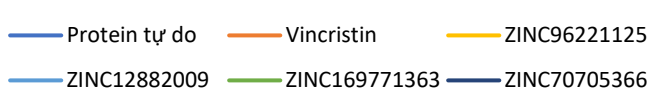


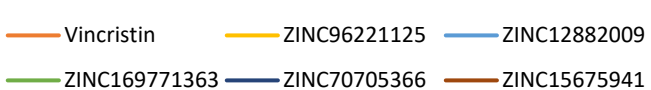
RMSD of protein



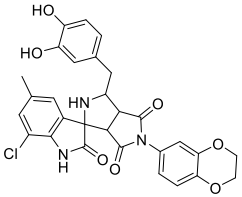
RMSD of ligand



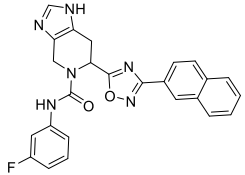




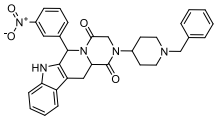
ZINC70705366 (D)
Docking score: -11.313 kcal/mol



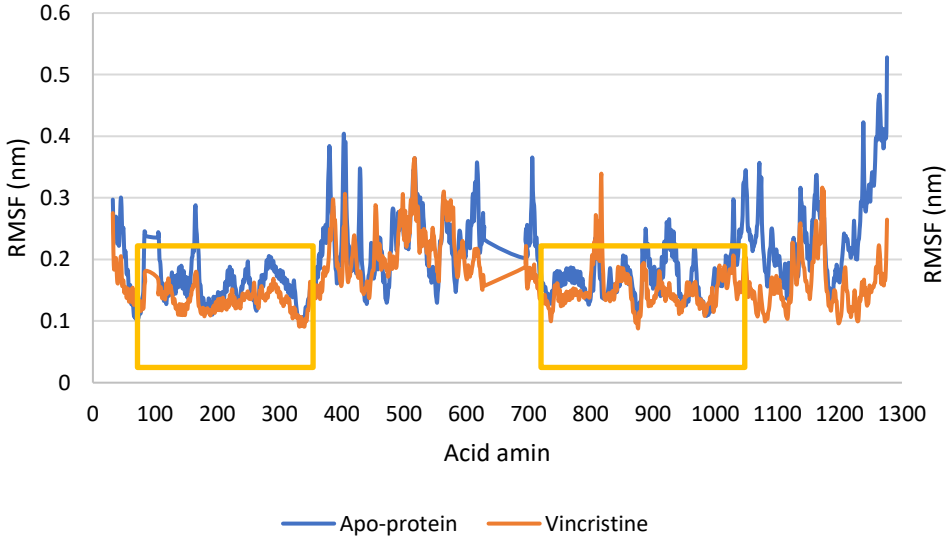
ZINC15675941 (E)
Docking score: -11.310 kcal/mol

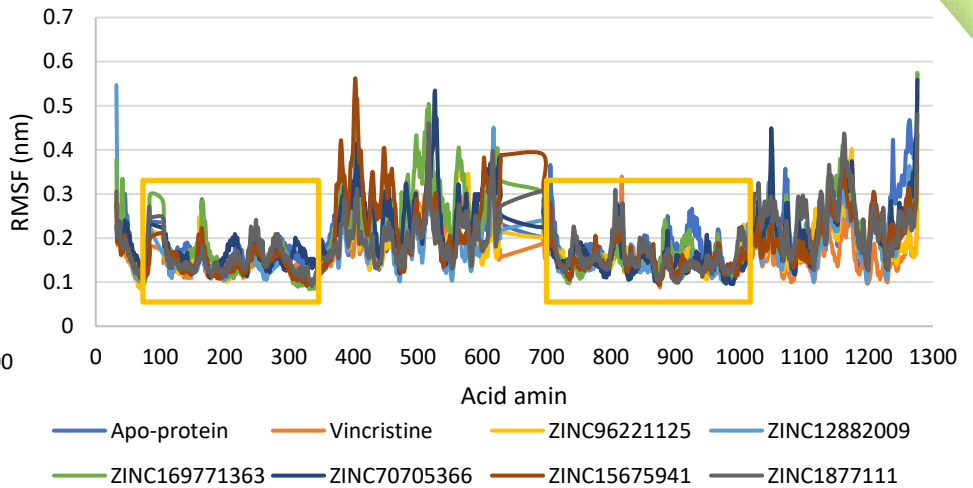


ZINC1877111 (F)
Docking score: -11.307 kcal/mol

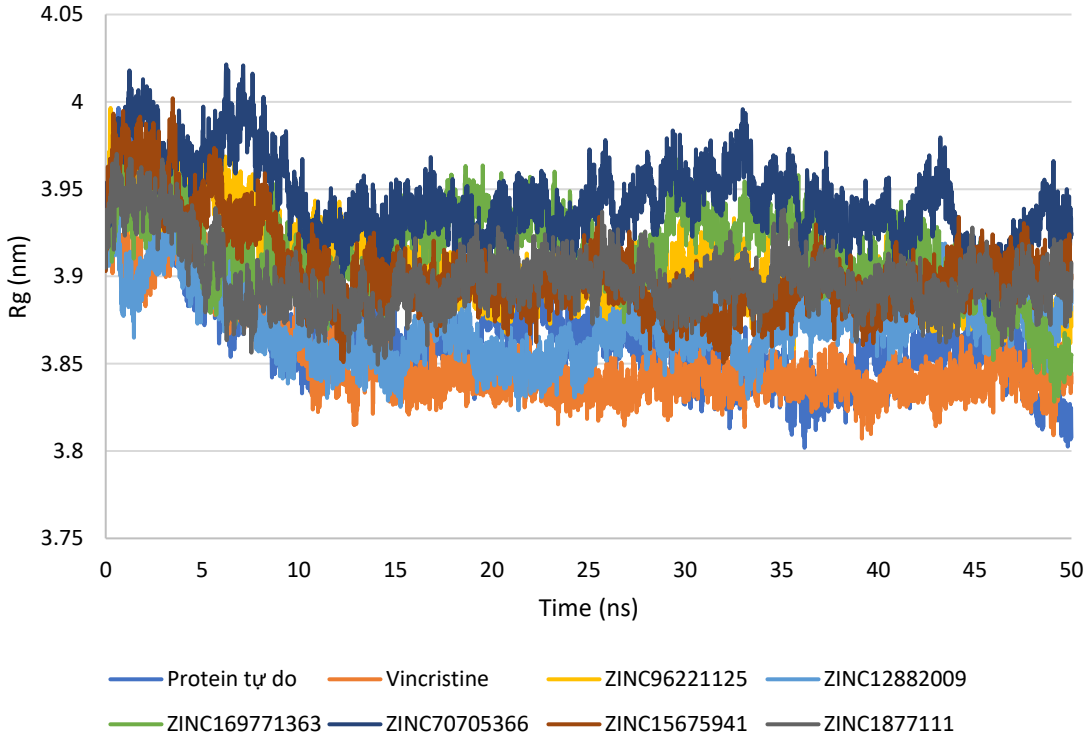


RMSF – Root Mean Square Fluctuation

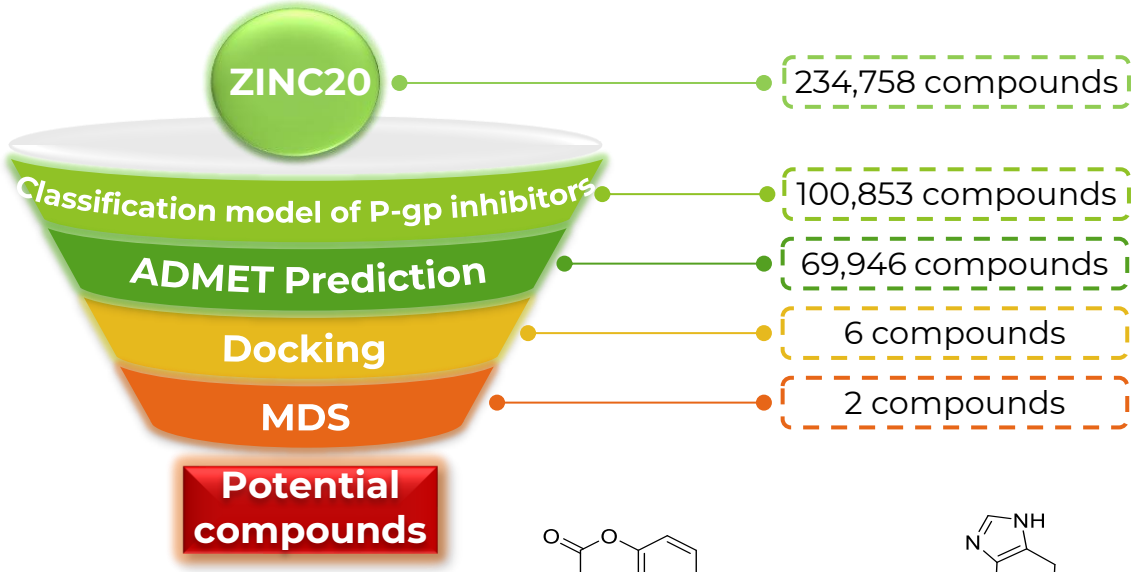


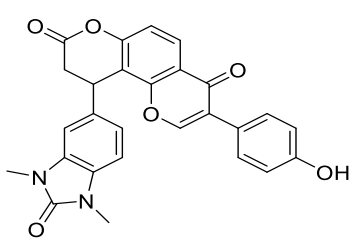


Rg – Radius of gyration

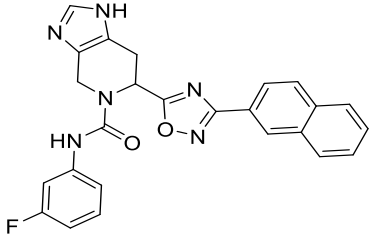


Result of screening P-gp inhibitors





ZINC96221125 (A)



ZINC15675941 (E)

Result of evaluating other datasets

Test sets	Classification threshold	TP	FP	TN	FN	Accuracy	Precision	Recall	F1-score	MCC
This research	Inhibitor: IC50 ≤ 15 µm Non-inhibitor: IC50 > 15 µm	147	49	91	15	0.79	0.75	0.91	0.82	0.58
ChEMBL (10%)	Inhibitor: IC50 ≤ 15 µm Non-inhibitor: IC50 > 15 µm	67	14	56	3	0.88	0.83	0.96	0.89	0.77
GF.Ecker et al	Inhibitor: IC50 ≤ 15 µm Non-inhibitor: IC50 > 100 µm	864	220	406	338	0.69	0.80	0.72	0.76	0.35
B.Zdrazil et al	Inhibitor: IC50 ≤ 15 µm Non-inhibitor: IC50 > 100 µm	66	59	57	9	0.64	0.53	0.88	0.66	0.38

Suggestion

- Building model classification based on the ability of inhibitors: Strong - Weak - Non-inhibit.
- Using a standardized method to input data.

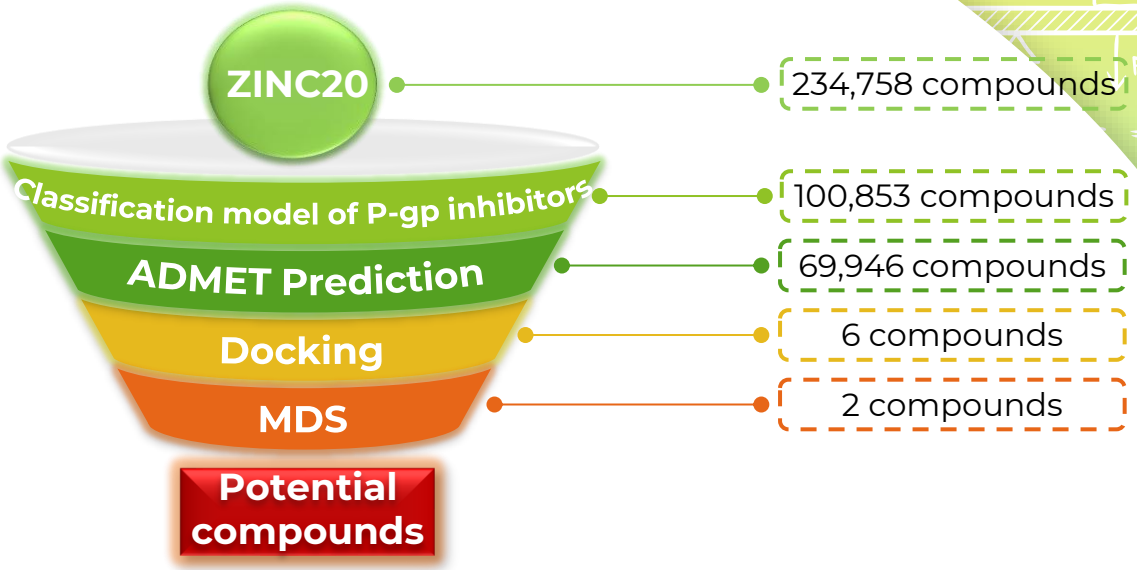
Classification models were published

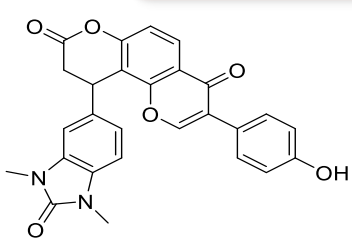
Research	Year	Methods	Databases	Accuracy
H.Sun et al	2005	NB	609 compounds	82.2%
YH.Wang et al	2005	Kohonen	206 compounds	82.3%
		BPNN	174 compounds	29%
P.Crivori et al	2006	PLSD	148 compounds	82%
F.Broccatelli et al	2011	MIF	1195 compounds	86%
L.Chen et al	2011	DT, RF, NB	1273 compounds	81.2%
S.Rapposelli et al	2012	RF, C4.5, consensus	59 compounds	75% (RF)
V.Poongavanam et al	2012	RF, kNN, SVM	1935 compounds	75% (RF)
F.Klepsch et al	2014	kNN, SVM, RF, DT, Binary QSAR	1954 compounds	73% (RF)

Research	Year	Methods	Databases	Accuracy
S.Eríc et al	2014	ANN, SVM, combine	135 compounds	89%
Khac-Minh Thai et al	2015	Backpropagation neural network	133 compounds	82%
V.Prachayasittikul et al	2015	DT, SVM, ANN	2477 compounds	87,4% (ANN)
M.Yang et al	2015	18 models machine learning	2428 compounds	77.4-79.9%
VK.Hinge et al	2019	GBM, GLM, SVM, kNN	1274 compounds	87.1-96.9% (SVM)
Khac-Minh Thai et al	2016	BPNN, C5.0, C&R Tree, QUEST, CHAID, Logistic Regression, Bayesian Network, Discriminant Analysis, SVM, combine	2131 compounds	92% (validation set) 100% (test set)
This research	**2022**	**CNN**	**1404 compounds**	**79%**

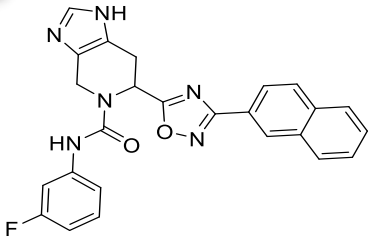
Conclusion

1. Building a deep learning model to classify P-gp inhibitors with good evaluation results
2. Predicted 100,853 natural compounds capable of inhibiting P-gp
3. Two potential substances to inhibit P-gp are ZINC96221125 (A), ZINC15675941 (E)





ZINC96221125 (A)



ZINC15675941 (E)

Suggestion

1. Extending simulation time to 100ns.
2. Testing *in vitro* and synthesizing (if necessary) potential compounds.
3. Building classification models for substrates and inhibitors.

Acknowledgement

Drug Discovery Laboratory
Department of Pharmaceutical Chemistry
Faculty of Pharmacy – UMP HCMC

Supervisor

Prof Dr Pharm Khac-Minh Thai
thaikhacminh@ump.edu.vn

Pharm Dac-Nhan Nguyen

Members

Pharm Thanh-Phuong Vo

Phuong-Uyen Tran-Thi

Pharm Thi-Thao-Nhung Nguyen

Minh Quang Mai

Cong-Thanh Phan-Van


Thank you for listening!