

Design and screening of HepG2 cancer cell line inhibitors from Triterpenoid derivatives of *Paramignya Trimera*

Tran Tu Uyen¹, Pham Van Tat², Nguyen Minh Quang^{3,*}



Use your smartphone to scan this QR code and download this article

¹Faculty of Pharmacy, Ho Chi Minh City University of Technology, 475A Dien Bien Phu Street, Binh Thanh District, Ho Chi Minh City, Viet Nam

²Institute of Pharmaceutical Education and Research, Binh Duong University, 504 Binh Duong Avenue, Thu Dau Mot City, Binh Duong, Vietnam

³Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City, 12 Nguyen Van Bao Street, Go Vap District, Ho Chi Minh City, Viet Nam

Correspondence

Nguyen Minh Quang, Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City, 12 Nguyen Van Bao Street, Go Vap District, Ho Chi Minh City, Viet Nam

Email: nguyenminhquang@iuh.edu.vn

History

- Received: 2023-10-04
- Accepted: 2023-11-16
- Published Online: 2023-6-30

DOI :

<https://doi.org/10.32508/stdj.v26iSI.4192>



ABSTRACT

Currently, artificial intelligence (AI) is a ubiquitous technology that provides effective support across all fields. In general, the pharmaceutical industry and drug production and development industry, in particular, are enjoying a very good application for the opportunity when *in silico* models have emerged as powerful platforms for designing new drugs. The aim of this project was to develop new anticancer agents by designing novel triterpenoid derivatives from *Paramignya Trimera* and predicting their efficacy against the Bcl-2 target receptor. The project used three main *in silico* models: QSAR_{MLR}, QSAR_{PCR} and QSAR_{ANN}. The models can be used to estimate IC₅₀ values for novel derivatives and for Escin extracted from *Paramignya Trimera*. Finally, the new good-value derivatives were docked to the Bcl-2 receptor to assess responsiveness. As a result, 196 newly designed compounds from the structural framework of the triterpenoid compounds were designed by combination with potential substituents. Screening by Veber identified 138 substances that met the requirement of having the ability to make drugs. Similarly, the QSAR_{MLR}, QSAR_{PCR}, and QSAR_{ANN} models were constructed according to the following statistical values: $R^2 = 0.849$, $R^2_{adj} = 0.826$, and $Q^2_{LOO} = 0.789$ for the QSAR_{MLR} model; the QSAR_{PCR} model, $R^2 = 0.860$, $R^2_{adj} = 0.831$, and $Q^2_{LOO} = 0.805$ and the QSAR_{ANN} model, $R^2_{train} = 0.941$, $R^2_{test} = 0.915$, and $R^2_{cv} = 0.912$. The use of models can help predict the effectiveness of newly engineered compounds. In the present study, 20 compounds were found to be more effective than Escin. Molecular docking on the Bcl-2 receptor revealed that T.new7 had the most potential, with a binding energy $E_{binding} = -7.933$ (kcal.mol⁻¹) and RMSD = 1.915 (Å). Research has achieved this goal by identifying T. new7, a newly designed compound with better anticancer efficacy than natural Escin.

Key words: Bcl-2, HepG2, Molecular docking, Paramignya trimer, QSAR, Triterpenoid

INTRODUCTION

According to GLOBOCAN, liver cancer is one of the 5 deadliest cancers, with a high number of new cases and deaths each year in 2020. Figure 1 shows that liver cancer has the third highest number of deaths in the world and the highest number of deaths in Vietnam¹. Worldwide, liver cancer is the third most common cause of cancer death, accounting for 8.3%, following lung and colorectal cancer. In Vietnam, liver cancer is the leading cause of death, accounting for 20.6% of all deaths. These data indicate that with the current situation of deaths from liver cancer, project implementation is extremely necessary. Liver cancer not only causes hundreds of thousands of deaths annually but also imposes a significant socioeconomic burden. Therefore, the search for liver cancer derivatives is extremely urgent for patients and for humanity in general.

Liver cancer is a type of cancer that starts in liver cells. The liver is a football-sized organ located in the upper right quadrant of the belly, beneath the diaphragm

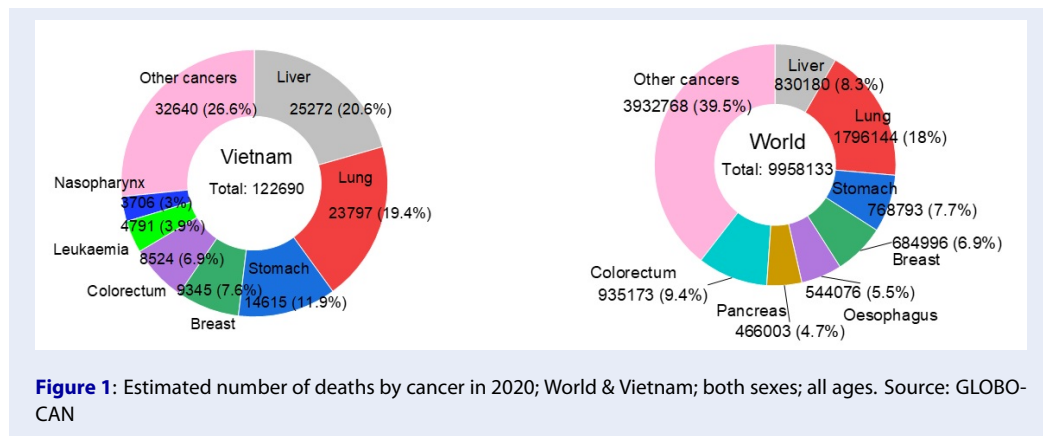
and above the stomach. The liver can develop several types of cancer. Hepatocellular carcinoma (HCC) is the most common type of liver cancer and is the main type of liver cell (hepatocyte). Cancer that spreads to the liver is more common than cancer that spreads to liver cells. Cancer that develops in another part of the body, such as the colon, lung, or breast, and spreads to the liver is referred to as metastatic cancer rather than liver cancer. This type of cancer is termed by the organ in which it began; for example, metastatic colon cancer describes cancer that begins in the colon and travels to the liver.

Triterpenes are a class of terpenes made up of six isoprene units with the chemical formula C₃₀H₄₈. These compounds can alternatively be thought of as three terpene units. Triterpenes are produced by animals, plants, and fungi and include squalene, the precursor to all steroids. Triterpenes have a wide range of structures. Almost 200 distinct skeletons have been identified. These skeletons can be roughly classified based on the number of rings present. Pentacyclic structures

Cite this article : Uyen T T, Tat P V, Quang N M. Design and screening of HepG2 cancer cell line inhibitors from Triterpenoid derivatives of *Paramignya Trimera*. *Sci. Tech. Dev. J.* 2024; 26(Special issue):1-16.

Copyright

© VNUHCM Press. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



(5 rings) predominate in general. One of the uses of Triterpenoids in the human body is to help prevent and treat cancer as well as to combat cancer metastasis. According to a 2011 study by Watchtel-Galor and colleagues, the use of triterpenoids in genital mushrooms has anticancer effects in vivo according to animal studies (mouse studies). In addition, the study indicated that the ingredients also contain active substances that help prevent cancer cells from growing in vitro (in the test tube). Thus, Triterpenoids help inhibit many types of cancer cells, such as lung cancer, breast cancer, and skin cancer cells. In addition, cancer metastasis is quite complicated. Cancer cells separate from the primary tumor and begin to move to other parts of the body. From there, small tumors—secondary tumors—form².

Triterpenoids are of interest because of their anti-inflammatory and analgesic properties, especially in anticancer cell lines, including HepG2 cells. Artificial intelligence facilitates the creation of virtual screening models for derivative compounds. This prospective study was designed to explore a synthetic compound with superior cancer-fighting properties compared to the natural substance found in *Paramignya trimera*. The triterpenoid of the oleanolic acid (OA) subgroup, called escin (Figure 3), is extracted from *Paramignya trimera* (Figure 2) of the family *Rutaceae*. OA has various benefits, including anti-inflammatory, antiviral, and hypoglycemic effects, and has potential for use against cancer cells. A large number of Triterpenoids are active against various human cancer cell lines, such as HepG2, SMMC-772 (hepatocellular carcinoma), HL-60 (leukemia), A549 (hepatocellular carcinoma), MCF-7 (breast cancer), and SW-480 (colon carcinoma)³.

Furthermore, oleanolic acid (OA) affects cancer cells via many routes. Increasing Bcl-2 receptor inhibition

is a strategy that promotes the proliferation of OA-treated HepG2 cancer cells. As a result, the Bcl-2 receptor was chosen as the target of interest. This research used *in silico* algorithms to predict novel synthetic chemicals. Chemicals that are more effective at inhibiting HepG2 cells are being found. The topic is focused on developing three models: QSAR_{MLR}, QSAR_{PCR}, and QSAR_{ANN}. Using virtual screening procedures saves time, money, and human resources. This research is likely to yield a result that speeds up compound screening in research and new medicine manufacture compared to experimental trials.

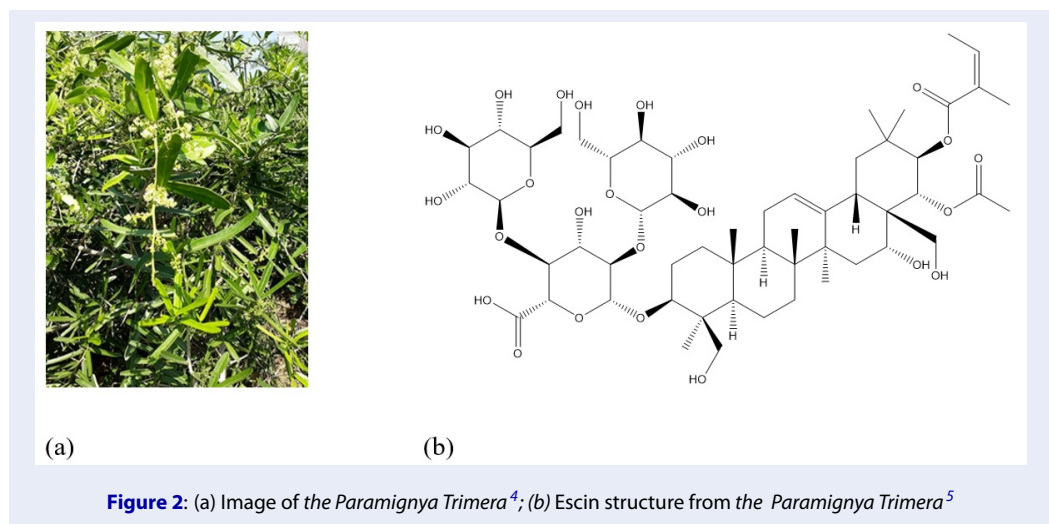
METHODOLOGY

Data mining from experiments

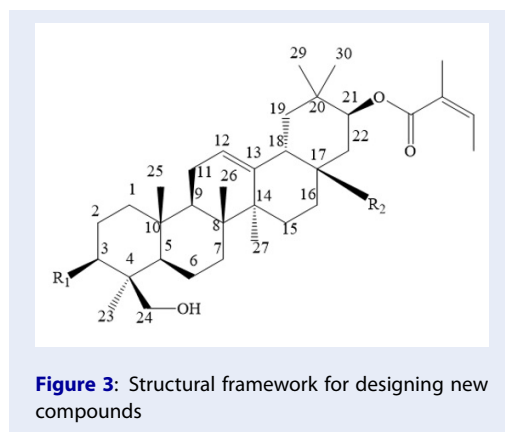
The data collected from the experiments are divided into 2 datasets: the training subset and the external evaluation subset. The two subsets are completely independent datasets. The condition is that the compound has a Triterpenoid framework and was tested on HepG2 cell carcinoma cells with an IC₅₀ value.

Design of new compounds

Two R1 and R2 sites (Figure 3) in the structural frame were selected for the attachment of substituents via the maximum design method. The binding group, which includes 14 cells labeled T1 to T14, has been shown to have anticancer activity. Therefore, 196 novel compounds were synthesized using the maximum design technique. Multilevel design: This method is used in drug design and helps generate a list of design compounds based on various taxonomic factors and material quantities⁶. The maximum design limits the possibility of missing significant compounds, thereby providing a complete dataset of possible designs when combining taxonomic elements (2 positions selected on the frame Escin structure) and



corresponding materials (14 functional groups T1-14 were selected).



Optimization of the structures

New derivatives were created using the ChemDraw program. Using molecular and quantum mechanics, all novel and experimental derivatives have been structurally optimized. These two types of software used include HyperChem with an MM+ force field and a gradient level of 0.05 and MOPAC with the semiempirical PM7 approach. This approach helps molecules determine the most stable structure and acquire descriptive variables, including partial charge, HOMO, LUMO, MW, DH, and so on.

Calculation of molecular descriptors

After the structural optimization procedure, MOE software was used to determine the molecular descriptor for all the datasets. The selection and cal-

culatation of all descriptors were investigated from 0 to 3D. When the results are made public. The variable screening procedure was used to exclude variables that were not important. When combined with the descriptor variables produced from the structural optimization process, a dataset of descriptive characteristics for each molecule is obtained, allowing QSAR models to be built.

Estimation of QSAR models

This study focused on the development of three QSAR models: multivariate linear regression (MLR), principal component regression (PCR), and artificial neural network (ANN) models.

QSAR_{MLR} model

The QSAR_{MLR} model predicts the dependent variable Y based on the values of two or more independent variables X. The model is represented as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e \quad (1)$$

Here, Y is the dependent variable, and $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ are the regression parameters of the model. X_k is the independent variable (k is the number of variables), and e is the random error. In this study, the dependent variable was the IC₅₀ value. The independent variable is the molecular descriptor²². Regression 2008²³ software was used to construct the QSAR_{MLR} model.

QSAR_{PCR} model

The set {X, Y}, where X is a data group with m observations and n variables and Y is the dependent variable. The information is gathered but not previously processed. Although outcome Y has no direct association

Symbol	Name	Structure	Symbol	Name	Structure
T1	Thiazol (7,8)		T8	Vanillin(9)	
T2	Pyrazole(10)		T9	Triazole(11)	
T3	Pyrimidin(8,12)		T10	Allopurinol(13)	
T4	Coumarin(14)		T11	Piperazine(15)	
T5	Quinoline(16)		T12	Morpholine(17)	
T6	Sulfamid(18)		T13	Mefenamic acid(19)	
T7	Imidazole(16,20)		T14	Indole(21)	

Figure 4: The substitution groups used in the design of new molecules⁷⁻²¹

with X, it does have a relationship with the principal components, which is a property of principal component regression²². To create the QSAR_{PCR} model, the XLSTAT 2016²⁴ program was used.

QSAR_{ANN} model

Artificial neural networks (ANNs) perform the same learning process as the human brain²². The structure of an artificial neural network I(m)-HL(n)-O(k) includes the following: the input layer I(m) is the descriptive variable of the QSAR_{MLR} model, the output layer O(k) is the IC₅₀ value, and the hidden layer HL(n) is investigated to determine the best QSAR_{ANN} model²⁵. The QSAR_{ANN} model was trained on the MATLAB 2016²⁶ tool.

Drug-likeness

The rule of action for Lipinski-5, the earliest and most well-known rule for identifying substances with good oral absorption, was proposed in 1997²⁷. Since then, several analogous rules based on molecular characteristics, such as those given by Ghose²⁸ and Veber²⁹, have been established. According to Veber's rule, substances in this study must meet the following two cri-

teria: rotatable bonds (nRB) ≤ 10 and polar surface area (tPSA) $\leq 140 \text{ \AA}^2$. Therefore, screening according to the same rules is aimed at finding compounds that have the potential to become more effective oral drugs according to Veber's rule under two conditions: rotatable bonds (nRB) ≤ 10 and polar surface area (tPSA) $\leq 140 \text{ \AA}^2$. Reduced molecular flexibility, as measured by the number of rotatable bonds and low polar surface area or total hydrogen bond count (sum of donors and acceptors), are important predictors of good oral bioavailability. A reduced polar surface area correlates better with an increased permeation rate than does lipophilicity (C log P), and an increased rotatable bond count has a negative effect on the permeation rate²⁹.

Bioactivity prediction

Medicinal characteristics were determined by using three QSAR models to predict the bioactivity at the IC₅₀ for new synthetic compounds, and esin is a natural triterpenoid derived from *Paramignya Trimera*. Then, using Escin as a reference, we looked for derivatives with greater biological activity than Escin. Currently, research predicts and discovers derivatives

with higher bioavailability than natural chemicals and the potential to become medications.

Molecular Docking

The main goal of molecular docking is to understand and predict molecular recognition both in terms of structure-finding bonds and energy-predicting affinity. Currently, the application of molecular docking methods is very diverse and includes structure-activity studies, optimization, and potential molecule searches via virtual screening³⁰. In this study, we used the MOE2019 package to perform the molecular docking process.

Escin, a Triterpenoid derived from *Paramignya Trimera*, belongs to the oleanolic acid group. This process affects cancer cells in a variety of ways, including inducing cyclic death, controlling the cell cycle, and killing cancer cells. Bcl-2 normally prevents cell cycle death (apoptosis). The target of action in this investigation was chosen to be Bcl-2, which inhibits Bcl-2 receptors, hence boosting cancer cell cyclic death³¹. The Bcl-2 receptor, encoded 4D2M, was obtained from the Protein Data Bank (PDB)³².

RESULTS

The training and test datasets

Seventy-four chemicals were gathered from articles published in reputable journals and PubMed. The data were utilized to develop models and assess the external inhibitory concentration (IC_{50}). The training set of 60 compounds was used to develop QSAR models, and the external validation set of 14 compounds was utilized to assess the predictive power of the biological activity of the model.

Design of the new compound

Using the multilayer design method and ChemDraw tool, a total of 196 novel molecules were obtained. All of these derivatives were optimized using the proper molecular mechanics sequence, followed by PM7 quantum mechanics. They are then calculated descriptors in the following phase.

Optimization of the structure and calculation of descriptors

All the compounds, both experimental and newly created, were subjected to structural optimization and molecular descriptor computations. The findings generated 310 descriptive attributes for each molecule utilized to construct the QSAR model.

Construction of the QSAR-MLR models

The model yields the following equation: $R^2 = 0.849$, $R^2_{adj} = 0.826$, and $Q^2_{LOO} = 0.789$:

$$IC_{50} (\mu M) = 3.739 + 2.247 \times x_1 + 85.94 \times x_2 + 24.36 \times x_3 + 3.13 + 0.156 \times x_4 + 5.24 \times x_5 - 0.03695 \times x_6 + 0.933 \times x_7 - 0.131 \times x_8 \quad (2)$$

Construction of the QSAR_{PCR} model

The QSAR_{PCR} model was built based on the variables of the QSAR_{MLR} model and yielded the following results: $R^2 = 0.860$, $R^2_{adj} = 0.831$, and $Q^2_{LOO} = 0.805$. The equations are represented as follows:

$$IC_{50} (\mu M) = 8.406 + 2.48 \times x_1 + 73.319 \times x_2 + 25.836 \times x_3 + 0.135 \times x_4 + 3.965 \times x_5 - 0.037 \times x_6 + 1.250 \times x_7 - 0.130 \times x_8 \quad (3)$$

Construction of the QSAR_{ANN} Models

The QSAR_{ANN} model was built using the QSAR-MLR model descriptors from equation (2). The training of models uses a back-propagation algorithm with transfer functions such as Logsig, Tansig and Purelin. Therefore, the architecture of the ANN models in this scenario is I(8)-HL(m)-O(1). QSAR_{ANN} models were developed in two stages. First, using the training dataset, multiple designs of MLP networks with different m values were identified, and the results are shown in Table 2.

Second, using the same external evaluation dataset (Table 5) of the QSAR_{MLR} and QSAR_{PCR} models, the best network was chosen based on the MARE (%) and Q^2_{EV} values. As a result, the best model was I(6)-HL(6)-O(1) (Figure 5a), with the statistical parameters $Q^2_{EV-ANN} = 0.866$ and MARE = 62.1%, as shown in Figure 5 and Figure 12, respectively, with the Purelin transfer function.

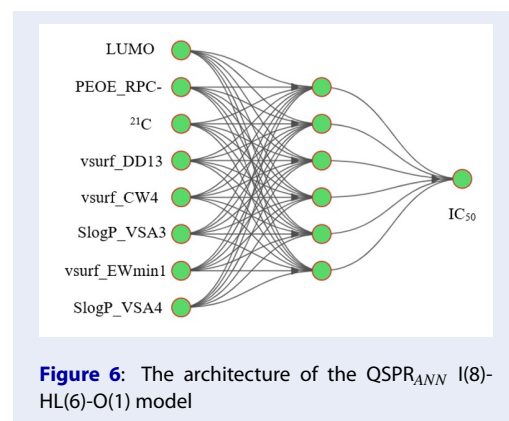


Figure 6: The architecture of the QSAR_{ANN} I(8)-HL(6)-O(1) model

Ord.	Structure	IC ₅₀ exp	Ord.	Structure	IC ₅₀ exp	Ord.	Structure	IC ₅₀ exp
1		1.00(33)	21		1.48(34)	41		10.50(35)
2		1.26(33)	22		0.90(14)	42		7.30(35)
3		6.26(33)	23		8.80(14)	43		10.20(35)
4		0.99(33)	24		3.20(36)	44		12.10(35)
5		1.48(33)	25		5.23(37)	45		0.85(38)
6		8.34(39)	26		15.20(40)	46		1.03(38)
7		4.88(39)	27		4.90(41)	47		0.90(38)
8		8.29(39)	28		7.50(41)	48		0.88(38)
9		9.34(39)	29		2.84(42)	49		0.97(38)
10		9.76(39)	30		2.22(42)	50		1.39(38)
11		7.40(38)	31		3.77(42)	51		2.47(38)
12		2.45(34)	32		2.19(38)	52		0.76(38)
13		4.03(34)	33		21.06(35)	53		1.02(38)
14		8.36(42)	34		22.78(35)	54		1.82(38)
15		1.90(43)	35		4.30(35)	55		4.65(38)
16		5.50(44)	36		6.50(35)	56		1.19(38)
17		0.60(44)	37		4.20(35)	57		0.91(38)
18		0.80(38)	38		5.20(35)	58		1.18(38)
19		3.90(38)	39		9.50(35)	59		1.00(38)
20		6.20(38)	40		7.30(35)	60		0.84(38)

Figure 5: Summary of the 60 experimental IC₅₀ values in the training datasets 33-47

Table 1: Results of building the QSAR_{MLR} model

k	Variables	R2	R2adj	Q2LOO	SE	Fstat	PRESS
1	x ₁	0.250	0.233	0.192	4.139	18.890	1064.026
2	x ₁ x ₂	0.626	0.613	0.569	2.938	47.792	568.351
3	x ₁ x ₂ x ₃	0.710	0.694	0.651	2.613	45.626	460.183
4	x ₁ x ₂ x ₃ x ₄	0.756	0.738	0.693	2.418	42.577	403.963
5	x ₁ x ₂ x ₃ x ₄ x ₅	0.785	0.765	0.720	2.291	39.363	368.313
6	x ₁ x ₂ x ₃ x ₄ x ₅ x ₆	0.814	0.793	0.752	2.150	38.628	326.307
7	x ₁ x ₂ x ₃ x ₄ x ₅ x ₆ x ₇	0.829	0.805	0.764	2.084	35.894	310.581
8	x ₁ x ₂ x ₃ x ₄ x ₅ x ₆ x ₇ x ₈	0.849	0.826	0.789	1.972	35.944	277.383
Notation of molecular descriptors							
LUMO	Lowest unoccupied molecular orbital	x ₁	vsurf_CW4		Capacity factor at -2.0		x5
PEOE_RF	Relative negative partial charge	x ₂	SlogP_VSA3		Bin 3 SlogP (0.00, 0.10]		x6
²¹ C	Partial charge of C position number 21	x ₃	vsurf_EWmin1		Lowest hydrophilic energy		x7
vsurf_DD	vsurf_EDmin1, vsurf_EDmin3 distance	x ₄	SlogP_VSA4		Bin 4 SlogP (0.10, 0.15]		x8

The external validation

External evaluation is considered a test to validate the predictive ability of the built models. From there, the model that gives the closest prediction results to the experimental results is selected. The external evaluation dataset includes 14 compounds obtained from experiments and is an independent set from the set used to construct the QSAR model. Detailed information on these derivatives is presented in Figure 7. The predicted values of the QSAR models for 14 substances in the external evaluation dataset are presented in Table 3.

Drug-likeness

After screening the drug likeness of all 196 compounds using Veber's criteria, we found 138 com-

pounds that met these criteria.

Bioactivity prediction

The study predicts the IC₅₀ for 138 newly designed substances and Escin. Then, 20 compounds with IC_{50s} less than Escin were obtained, ordered from small to large based on QSAR_{ANN} for the best predictability. The structures of the 20 potential substances are presented in Figure 8.

The detailed prediction results of 20 compounds from each model are presented in Table 4.

Molecular Docking

To test the inhibitory ability of the peptides on HepG2 cancer cells, 20 drugs were docked with the appropriate IC₅₀ values for the Bcl-2 receptor. This docking

Table 2: The results of initial screening for the ANN architecture

Ord.	QSARANN model	Transfer function	R2train	R2test	R2cv	Training error	Test error	Validation error	Training algorithm
1	I(8)-HL(6)-O(1)	Logsig	0.986	0.988	0.988	1.201	3.367	0.983	BFGS 42
2	I(8)-HL (6)-O(1)	Tansig	0.957	0.984	0.998	0.961	2.739	1.186	BFGS 63
3	I(8)-HL(10)-O(1)	Logsig	0.965	0.987	0.97	0.948	2.084	1.752	BFGS 39
4	I(8)-HL(10)-O(1)	Tansig	0.975	0.992	0.99	2.249	2.589	0.721	BFGS 13
5	I(8)-HL(5)-O(1)	Purelin	0.907	0.967	0.973	2.790	2.246	1.003	BFGS 17
6	I(8)-HL(6)-O(1)	Purelin	0.941	0.916	0.912	0.822	1.389	0.795	BFGS 36
7	I(8)-HL(10)-O(1)	Purelin	0.911	0.921	0.975	1.658	2.326	1.896	BFGS 43
8	I(8)-HL(4)-O(1)	Purelin	0.917	0.977	0.936	2.567	2.238	1.326	BFGS 69

Table 3: The predicted IC_{50,pred} values of the three models in the external evaluation set

Symbol	IC _{50,exp} (μM)	IC _{50,pred} (μM)		
		QSARMLR	QSARPCR	QSARANN6
TPN1	8.900	5.777	6.196	5.311
TPN2	29.700	10.388	10.573	11.069
TPN3	3.300	3.754	3.613	2.562
TPN4	17.660	6.408	6.462	7.848
TPN5	11.600	5.995	6.276	4.827
TPN6	15.500	8.868	9.065	7.713
TPN7	17.700	7.547	7.519	7.264
TPN8	22.400	8.867	9.065	7.714
TPN9	25.000	9.353	10.069	8.772
TPN10	23.700	7.547	7.519	7.264
TPN11	22.300	8.485	9.184	7.403
TPN12	21.000	10.116	10.320	10.313
TPN13	26.700	9.860	10.266	10.054
TPN14	0.660	3.750	4.015	1.5943

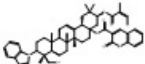
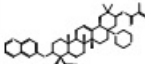
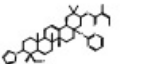
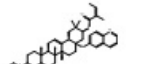
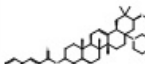
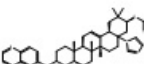
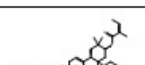
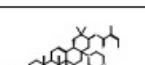
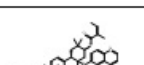
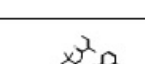
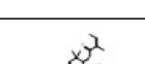
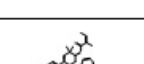
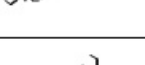


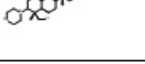
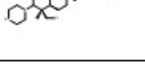
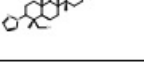
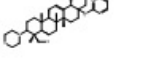
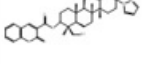
Symbol	Structure	Symbol	Structure	Symbol	Structure
T.new1		T.new8		T.new15	
T.new2		T.new9		T.new16	
T.new3		T.new10		T.new17	
T.new4		T.new11		T.new18	
T.new5		T.new12		T.new19	
T.new6		T.new13		T.new20	
T.new7		T.new14			

Figure 7: The experimental values of $IC_{50,exp}$ in the external evaluation dataset

Table 4: The predicted $IC_{50,pred}$ values of new derivatives from three QSAR models

Symbol	$IC_{50,pre} (\mu M)$			Symbol	$IC_{50,pre} (\mu M)$		
	QSAR _{MLR}	QSAR _{PCR}	QSAR _{ANN}		QSAR _{MLR}	QSAR _{PCR}	QSAR _{ANN}
T.new1	1.754	2.15	2.675	T.new11	4.025	4.277	1.817
T.new2	2.848	2.7	1.532	T.new12	5.571	5.589	2.502
T.new3	2.905	4.159	1.477	T.new13	4.11	4.786	1.672
T.new4	0.807	0.973	2.094	T.new14	3.906	4.175	1.955
T.new5	5.013	5.616	2.254	T.new15	5.042	5.325	2.201
T.new6	4.48	4.687	2.056	T.new16	1.229	2.004	2.279
T.new7	0.938	1.764	1.187	T.new17	5.299	5.871	2.686
T.new8	6.115	6.803	2.719	T.new18	2.863	3.642	1.573
T.new9	2.369	3.11	1.496	T.new19	1.818	2.746	1.306
T.new10	6.277	7.156	1.356	T.new20	3.685	4.609	1.957
				Escin	3.391	3.534	2.752

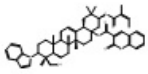
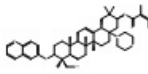
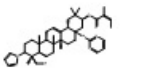
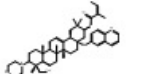
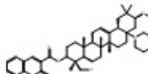
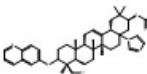
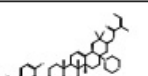
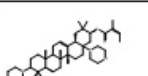
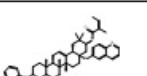
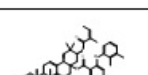
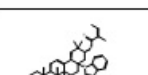
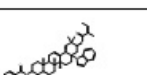
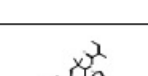
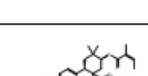
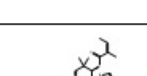
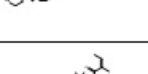

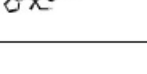
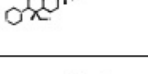
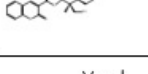
Symbol	Structure	Symbol	Structure	Symbol	Structure
T.new1		T.new8		T.new15	
T.new2		T.new9		T.new16	
T.new3		T.new10		T.new17	
T.new4		T.new11		T.new18	
T.new5		T.new12		T.new19	
T.new6		T.new13		T.new20	
T.new7		T.new14			

Figure 8: The structures of the 20 new compounds

process helps evaluate the binding ability of the compound to the Bcl-2 target receptor by simulating the 3D structure of both the receptor and the compound. Substances that are considered well bound have an RMSD < 2.0 Å and an E_{binding} < -7.0 kcal.mol⁻¹. The results for the 6 compounds with good binding energies and RMSD values are presented in Figure 5.

DISCUSSION

QSAR models

Table 1 and Figure 10 show that the R², R²_{adj}, and Q²_{LOO} values are proportional to the number of variables. This shows that when the number of variables increases, the model improves. This change was accompanied by a significant change from 7 to 8 variables, although before each increase, the variables did not change much. Therefore, 8 variables are needed, indicating that this is the most promising QSAR_{MLR} model. The model consists of 8 variables as follows: LUMO, PEOE_RPC- ²¹C vsurf_DD13 vsurf_CW4

SlogP_VSA3 vsurf_EWmin1 and SlogP_VSA4. LUMO is the lowest unoccupied molecular orbital; PEOE_RPC- is a relatively negative partial charge; ²¹C is the partial charge at position 21; vsurf_DD13 is the vsurf_EDmin1 - vsurf_EDmin3 distance; vsurf_CW4 is the capacity factor at -2.0; logP_VSA3 is the bin 3 SlogP (0.00, 0.10]; vsurf_EWmin1 is the lowest hydrophilic energy; and SlogP_VSA4 is the bin 4 SlogP (0.10, 0.15].

The QSAR_{MLR} model results for R² = 0.849 > 0.6(47) showed that the model encoded 84.9% of the biological activity variables in the dataset. An R²_{adj} = 0.826 > 0.5 represents an encoding of 82.6% of the active value variable in the data, and Q²_{LOO} = 0.789 > 0.5(47). As a result of these findings, the model produced relatively good prediction outcomes.

Moreover, the QSAR_{PCR} model results for R² = 0.860 > 0.6(47) demonstrated that the model encoded 86% of the biological activity variables in the dataset. An R²_{adj} = 0.831 > 0.5 represents an encoding of 83.1%

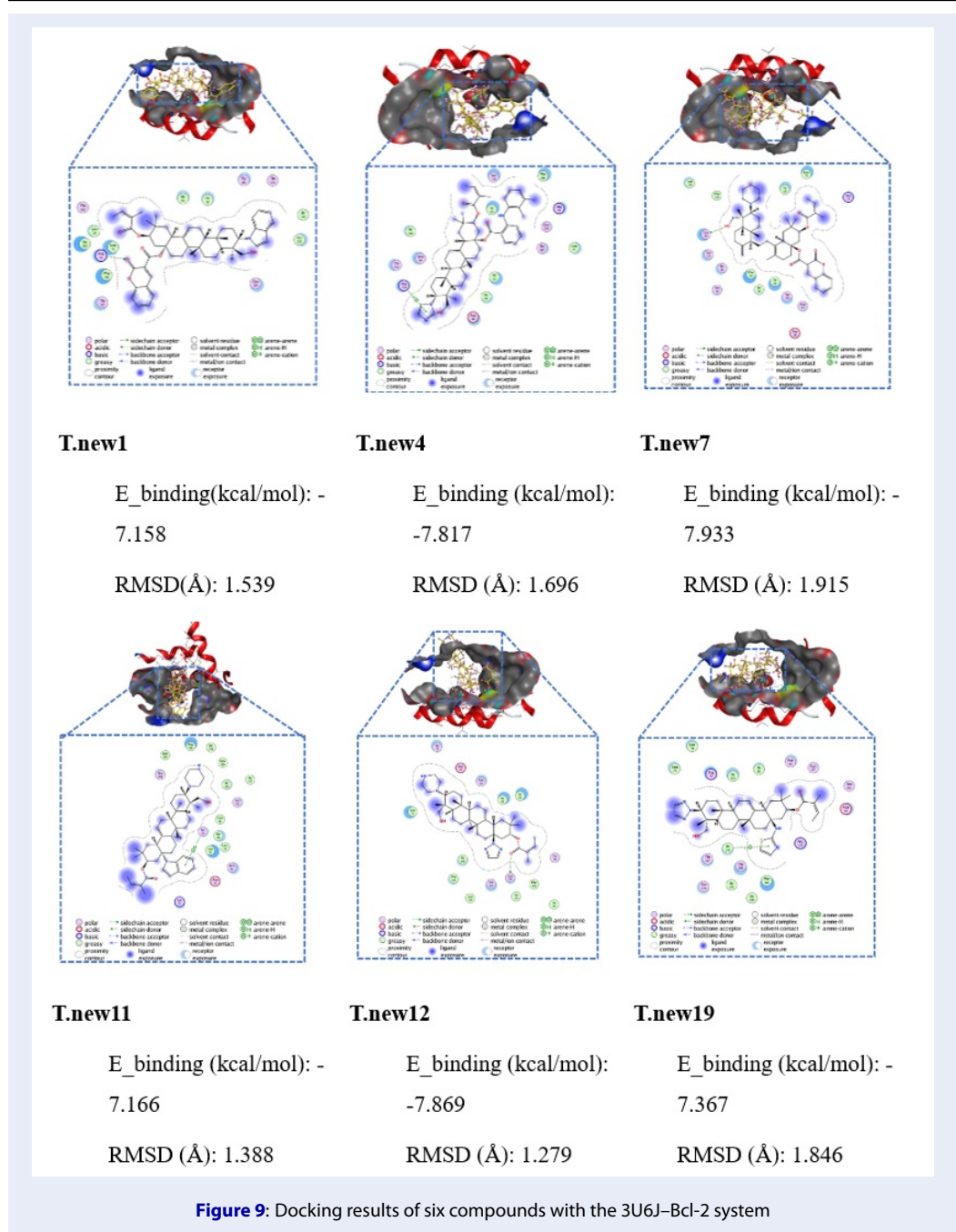


Figure 9: Docking results of six compounds with the 3U6J-Bcl-2 system

of the variable to the active value in the data. $Q^2_{LOO} = 0.846 > 0.5(47)$. Based on these findings, the model produces accurate predictions.

The architecture of ANN I(8)-HL(6)-O(1) using the Purelin transfer function for $R^2_{train} = 0.941$, $R^2_{test} = 0.916$ and $Q^2_{cv} = 0.912$ shows that the model has good predictability with high correlation values. The results with an external evaluation set of 0.866 show that the predictive ability of this model is closest to reality.

Based on the above reasons, the QSAR models were chosen to develop the new design and Escin.

The contributions of the variables in the model were also investigated, and the results are presented in Figure 11. All the descriptors contributed significantly to various degrees the most significant contributor was PEOE_RPC-, and the least significant contributor was vsurf_EWmin1, with contributions of 42.3% and 1.6%, respectively. The remaining variables also

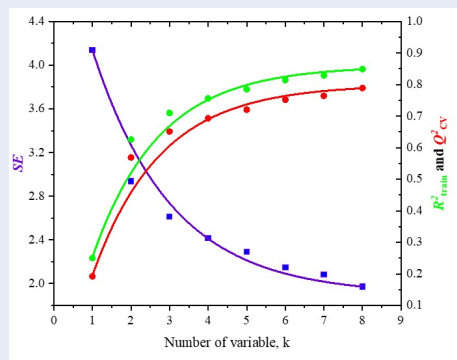


Figure 10: The variation in the SE, R^2_{train} , and Q^2_{LOO} values in response to the k value

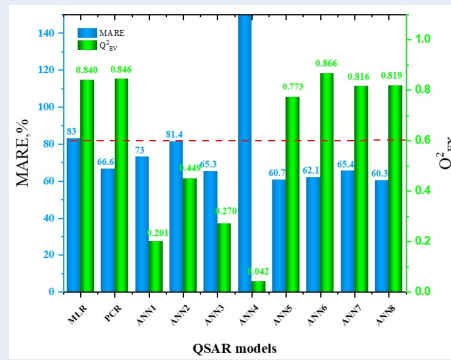


Figure 12: The MARE (%) and Q^2_{EX} values of the QSAR models

contributed to the $QSAR_{MLR}$ model in the following order: $PEOE_{RPC} > vsurf_{CW4} > SlogP_{VSA3} > LUMO > 21C > vsurf_{DD13} > SlogP_{VSA4} > vsurf_{EWmin1}$.

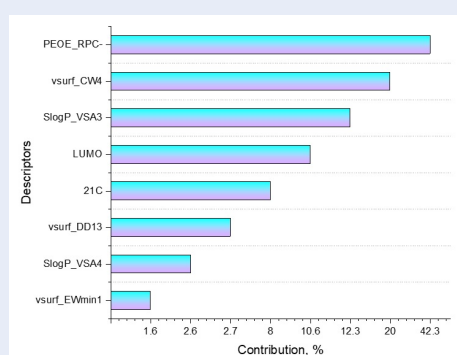


Figure 11: The contributions of the variables to the $QSAR_{MLR}$ model

The external validation

As mentioned above, external evaluation is used to construct the MLR and PCR models. In addition, the best ANN model was identified from the initial survey models, as shown in Table 2. The two values used as a basis for evaluation are Q^2_{EX} (>0.5) and MARE (%). The results are fully presented in Figure 12. The results show that the linear regression models meet the requirements, and the 6th neural network model (ANN6) is selected for the project because the Q^2_{EX} value is 0.866, which is the highest, while the MARE (%) value is comparable to that of the other models. As depicted in Figure 13, the Q^2_{EX} values of the $QSAR_{MLR}$ model for the relationships between the $IC_{50,pred}$ and experimental $IC_{50,exp}$ values are shown,

for a value of 0.840. Similar to the $QSAR_{PCR}$ and the external evaluation set, the result is $Q^2_{EX} = 0.846$, and the $QSAR_{ANN6}$ gives a result of 0.866. The conclusion that the above three models all give good correlation index results for the external evaluation set shows that the evaluation ability is reliable and can be used to predict a wide range of design substances.

Furthermore, one-way ANOVA showed that the differences between the experimental and predicted values from the three models, $QSAR_{MLR}$, $QSAR_{PCR}$, and $QSAR_{ANN}$, were not significant when the results were $F = 0.0269 < F_{0.05} = 3.2381$. Therefore, the predictive ability of the three models is appropriate.

Bioactivity prediction

Under the same calculation conditions, the same models predicting the results obtained above for the 20 substances had better IC_{50} values than did those of Escin. The present study used Escin as a base to select compounds with better biological activity to prove that this potential new substance has superior cancer cell inhibitory ability compared to natural active substances.

Prediction results of new molecules and predicted Escin values from three QSAR models, $QSAR_{MLR}$, $QSAR_{PCR}$, and $QSAR_{ANN6}$. There was no significant difference in the analysis of variance ($F = 0.71595 < F_{0.05} = 3.07606$). Therefore, the predictive ability of the three models is consistent and reliable.

Molecular Docking

The full interaction results of the six new compounds on the Bcl-2 receptor are presented in Table 5. Among the six compounds that gave good results

T.new1 binds to the Bcl-2 receptor via a hydrogen acceptor bond to ARG74 (distance = 2.97 Å, energy =

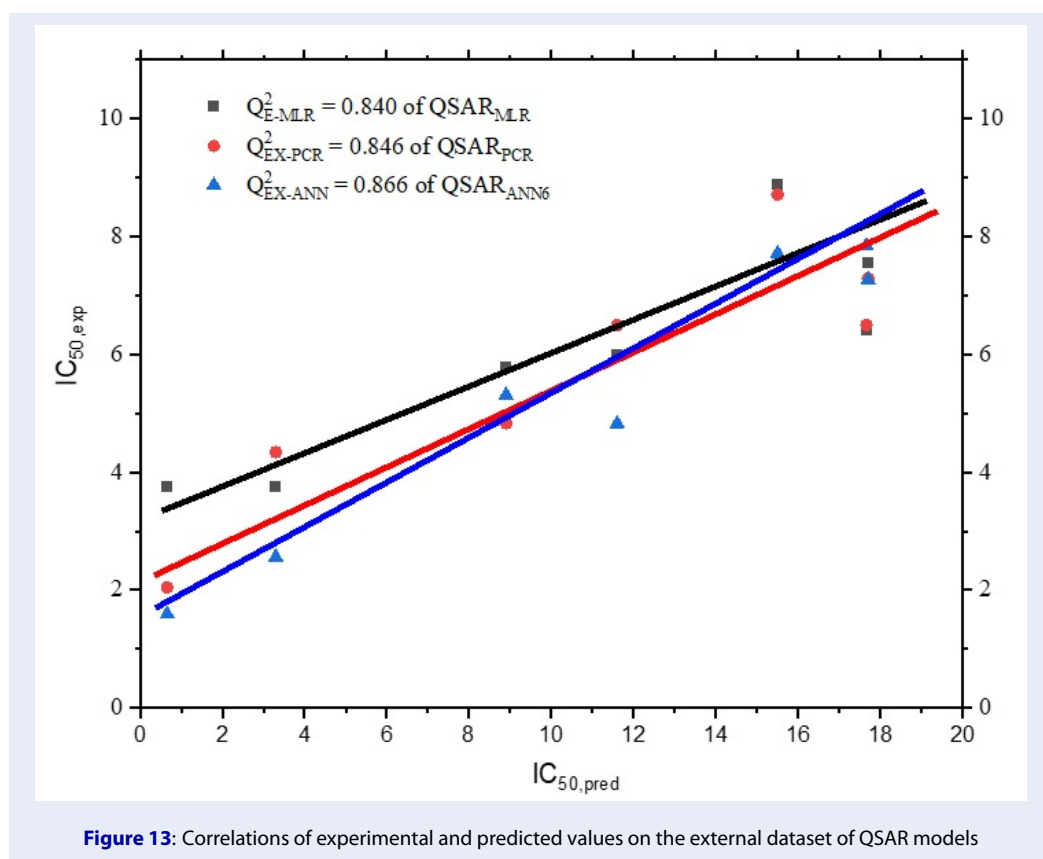


Figure 13: Correlations of experimental and predicted values on the external dataset of QSAR models

-1.7 kcalmol⁻¹), $E_{binding} = -7.158$ (kcalmol⁻¹) and RMSD = 1.539 (Å).

T.new4 binds to the Bcl-2 receptor via a pi-cation bond to ARG154 (distance = 3.58 Å, energy = -1.6 kcalmol⁻¹), $E_{binding} = -7.817$ (kcalmol⁻¹) and RMSD = 1.696 (Å).

T.new7 binds to the Bcl-2 receptor via a hydrogen donor bond to CYS174 (distance = 3.37 Å, energy = -1.2 kcalmol⁻¹), $E_{binding} = -7.933$ (kcalmol⁻¹) and RMSD = 1.915 (Å).

T.new11 binds to the Bcl-2 receptor via a pi-cation bond to TYR79 (distance = 3.78 Å, energy = -0.9 kcalmol⁻¹), $E_{binding} = -7.166$ (kcalmol⁻¹) and RMSD = 1.388 (Å).

T.new12 binds to the Bcl-2 receptor via a hydrogen donor bond to CYS174 (distance = 3.23 Å, energy = -0.8 kcalmol⁻¹), $E_{binding} = -7.869$ (kcalmol⁻¹) and RMSD = 1.279 (Å).

T.new19 binds to the Bcl-2 receptor via a pi-cation bond to ILE146 (distance = 3.65 Å, energy = -0.9 kcalmol⁻¹), $E_{binding} = -7.367$ (kcalmol⁻¹) and RMSD = 1.846 (Å).

The full interaction results of the six new compounds on the Bcl-2 receptor are presented in Table 5. Among

the six compounds that gave good results, T.new7 had the best results: T.new7 binds to the Bcl-2 receptor by a hydrogen donor bond to CYS174 (distance = 3.37 Å, energy = -1.2 kcalmol⁻¹), $E_{binding} = -7.933$ (kcalmol⁻¹) and RMSD = 1.915 (Å).

T.new7 had the best results: T.new7 binds to the Bcl-2 receptor by a hydrogen donor bond to CYS174 (distance = 3.37 Å, energy = -1.2 kcalmol⁻¹) $E_{binding} = -7.933$ (kcalmol⁻¹) and RMSD = 1.915 (Å).

On the BCL-2 receptor, the amino acids considered essential are ARG74, ARG154, CYS174, TYR79, CYS174, and ILE146 when sequentially linked by the 6 most potential compounds.

Since then, $QSAR_{MLR}$, $QSAR_{PCR}$, and $QSAR_{ANN6}$ models have been successfully constructed to predict new engineered substances. Finally, the T.new7 compound was selected to inhibit HepG2 cancer cells.

CONCLUSION

This study applied the QSAR model to screen and develop new drugs, specifically triterpenoids, for use on HepG2 cancer cells. The final selected compound T.new7 showed better bioavailability than the naturally occurring substance and met the conditions

Table 5: Detailed interaction results of new compounds on the Bcl-2 receptor

Compound	Ligand	Receptor		Interaction	Distance (Å)	E (kcal/mol)	
T.new1	O 58	NE	ARG	74 (A)	H-acceptor	2.94	-1.7
T.new4	5-ring	NH2	ARG	154 (A)	pi-cation	3.58	-1.6
T.new7	O 20	SG	CYS	174 (A)	H-donor	3.37	-1.2
T.new11	6-ring	CD1	TYR	79 (A)	pi-H	3.78	-0.9
T.new12	O 48	SG	CYS	174 (A)	H-donor	3.23	-0.8
T.new19	5-ring	CG2	ILE	146 (A)	pi-H	3.65	-0.9

for use as a drug according to Veber's rule. Biological activity prediction based on the screening process and statistical statistics is fair and reliable. This study provides the foundation for future T.new7 experimental studies. Based on these findings, T.new7 was created using the Escin structural framework, with R₁ as morpholine and R₂ as coumarin. T.new7 has an IC₅₀ (μM) of 0.938, 1.764, and 1.187 according to the 3 models QSARMLR, QSARPCR, and QSARANN, respectively. In this case, QSAR_{ANN} produces the best prediction results, and all three values are lower than those of the natural parent chemical Escin. The results of molecular docking showed that E_{binding} = -7.933 kcal.mol⁻¹ and RMSD = 1.915 Å. The T.new7 compound binds to the Bcl-2 receptor via an H-donor. Specifically, T.new7 gives amino acid CYS174 a hydrogen (distance d = 3.37 Å, energy = -1.2 kcal.mol⁻¹). Therefore, T.new7 was selected as the best inhibitor of HepG2 cancer cells. This current study is limited by the fact that it involved only virtual screening. Despite its exploratory nature, this study offers T.new7 for further experiments. Consequently, further experimental studies are necessary to confirm the effectiveness of T.new7 against HepG2 cancer cells.

ABBREVIATIONS

ANOVA: Analysis of variance

ANN: Artificial neural network

Bcl-2: B-cell lymphoma 2

GLOBOCAN: Global Cancer Statistics

HepG2: Human liver cancer cell line

HOMO: Highest occupied molecular orbital

IC₅₀: Half-maximal inhibitory concentration

LUMO: lowest unoccupied molecular orbital

MLR: Multiple linear regression

OA: Oleanolic Acid

PCR: Principal component regression

PDB: Protein Data Bank

PM7: Parameterized Model 7

QSAR: Quantitative structure-activity relationship

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

All authors participated in study design, coordination, and manuscript drafting

FUNDING

None

REFERENCES

1. Cancer today [Internet]. [cited 2023 Jun 25]; Available from: <http://gco.iarc.fr/today/home>.
2. Trigos Á, Medellín JS. Biologically active metabolites of the genus Ganoderma: Three decades of myco-chemistry research. 2011;
3. Chudzik M, Korzonek-Szlacheta I, Król W. Triterpenes as Potentially Cytotoxic Compounds. *Molecules*. 2015 Jan 19;20(1):1610-25; PMID: 25608043. Available from: <https://doi.org/10.3390/molecules20011610>.
4. Phi TCM, Chu HH, Trieu Le N, Nguyen DB. Phylogenetic relationship of Paramignya trimeria and its relatives: an evidence for the wide sexual compatibility. *Sci Rep*. 2020 Dec 10;10(1):21662; PMID: 33303786. Available from: <https://doi.org/10.1038/s41598-020-78448-2>.
5. Nguyen VT, Sakoff JA, Scarlett CJ. Physicochemical Properties, Antioxidant and Anti-proliferative Capacities of Dried Leaf and Its Extract from Xao tam phan (Paramignya trimeria). *Chem Biodivers*. 2017 Jun;14(6):e1600498; PMID: 28122160. Available from: <https://doi.org/10.1002/cbdv.201600498>.
6. Multilevel Categoric. Tutorials. 2021;
7. Jain S, Pattnaik S, Pathak K, Kumar S, Pathak D, Jain S, et al. Anticancer Potential of Thiazole Derivatives: A Retrospective Review. *Mini-Rev Med Chem*. 2018 Apr 16;18(8):640-55; PMID: 29173166. Available from: <https://doi.org/10.2174/1389557517666171123211321>.
8. Urban M, Vlk M, Dzubak P, Hajdich M, Sarek J. Cytotoxic heterocyclic triterpenoids derived from betulin and betulinic acid. *Bioorg Med Chem*. 2012 Jun;20(11):3666-74; PMID: 22551630. Available from: <https://doi.org/10.1016/j.bmc.2012.03.066>.
9. Al-Naqeb G, Ismail M, Bagalkotkar G, Adamu HA. Vanillin rich fraction regulates LDLR and HMGCR gene expression in HepG2 cells. *Food Res Int*. 2010 Dec;43(10):2437-43; Available from: <https://doi.org/10.1016/j.foodres.2010.09.015>.
10. Bennani FE, Doudach L, Karrouchi K, El Rhayam Y, Rudd CE, Ansar M, et al. Design and prediction of novel pyrazole derivatives as potential anticancer compounds based on 2D-QSAR study against PC-3, B16F10, K562, MDA-MB-231, A2780, ACHN and NUGC cancer cell lines. *Heliyon*. 2022

- Aug;8(8):e10003;PMID: 35965973. Available from: <https://doi.org/10.1016/j.heliyon.2022.e10003>.
21. Alam MM. 1,2,3-Triazole hybrids as anticancer agents: A review. *Arch Pharm (Weinheim)*. 2022 Jan;355(1):2100158;PMID: 34559414. Available from: <https://doi.org/10.1002/ardp.202100158>.
 22. Kang X, Hu J, Gao Z, Ju Y, Xu C. Synthesis, anti-proliferative and proapoptotic activity of novel oleanolic acid azaheterocyclic derivatives. *MedChemComm*. 2012;3(10):1245; Available from: <https://doi.org/10.1039/c2md20051a>.
 23. Li Y, Cao TT, Guo S, Zhong Q, Li CH, Li Y, et al. Discovery of Novel Allopurinol Derivatives with Anticancer Activity and Attenuated Xanthine Oxidase Inhibition. *Molecules*. 2016 Jun 20;21(6):771;PMID: 27331805. Available from: <https://doi.org/10.3390/molecules21060771>.
 24. Vega-Granados K, Medina-O'Donnell M, Rivas F, Reyes-Zurita FJ, Martinez A, Alvarez De Cienfuegos L, et al. Synthesis and Biological Activity of Triterpene-Coumarin Conjugates. *J Nat Prod*. 2021 May 28;84(5):1587-97;PMID: 33956447. Available from: <https://doi.org/10.1021/acs.jnatprod.1c00128>.
 25. Al-Ghorbani M, Gouda MA, Baashen M, Alharbi O, Almalki FA, Ranganatha LV. Piperazine Heterocycles as Potential Anticancer Agents: A Review. *Pharm Chem J*. 2022 Apr;56(1):29-37; Available from: <https://doi.org/10.1007/s11094-022-02597-z>.
 26. Borková ML. Triterpenoids With Anticancer Properties and Their mechanism of Action;
 27. Arshad F, Khan MF, Akhtar W, Alam MM, Nainwal LM, Kaushik SK, et al. Revealing quinquennial anticancer journey of morpholine: A SAR based review. *Eur J Med Chem*. 2019 Apr;167:324-56;PMID: 30776694. Available from: <https://doi.org/10.1016/j.ejmech.2019.02.015>.
 28. Casini A, Scozzafava A, Mastrolorenzo A, Supuran C. Sulfonamides and Sulfonylated Derivatives as Anticancer Agents. *Curr Cancer Drug Targets*. 2002 Mar 1;2(1):55-75;PMID: 12188921. Available from: <https://doi.org/10.2174/1568009023334060>.
 29. Patel SS, Tripathi R, Chavda VK, Savjani JK. Anticancer Potential of Mefenamic Acid Derivatives with Platelet-Derived Growth Factor Inhibitory Property. *Anticancer Agents Med Chem*. 2020 Jul 24;20(8):998-1008;PMID: 32294047. Available from: <https://doi.org/10.2174/187152062066200415100614>.
 30. Sharma P, LaRosa C, Antwi J, Govindarajan R, Werbovetz KA. Imidazoles as Potential Anticancer Agents: An Update on Recent Studies. *Molecules*. 2021 Jul 11;26(14):4213;PMID: 34299488. Available from: <https://doi.org/10.3390/molecules26144213>.
 31. Dadashpour S, Emami S. Indole in the target-based design of anticancer agents: A versatile scaffold with diverse mechanisms. *Eur J Med Chem*. 2018 Apr;150:9-29;PMID: 29505935. Available from: <https://doi.org/10.1016/j.ejmech.2018.02.065>.
 32. Roy K, Kar S, Das RN. QSAR/QSPR Modeling: Introduction. In: *A Primer on QSAR/QSPR Modeling* [Internet]. Cham: Springer International Publishing; 2015 [cited 2023 Sep 21]. p. 1-36. (SpringerBriefs in Molecular Science); Available from: https://doi.org/10.1007/978-3-319-17281-1_1.
 33. Steppan D.D . Werner J, Yeater P.R. Essential Regression and Experimental Design for Chemists and Engineers, Free Software Package, 1998. 1998; Available from: <http://home.t-online.de/home/jowerner98/indexeng.html>.
 34. XLSTAT. version 2016.02.28451. 2016; Available from: Addinsoft, USA;
 35. Rojas R. A Systematic Introduction. *Neural Networks*. 1996; Available from: <https://doi.org/10.1007/978-3-642-61068-4>.
 36. MATLAB. R2016a 9.0.0.341360. [Internet]. 2016. Available from: MathWorks, USA;
 37. Lipinski C, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 1997 Jan 1;3:23-5; Available from: [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1).
 38. Ghose AK, Viswanadhan VN, Wendoloski JJ. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J Comb Chem*. 1999 Jan 12;1(1):55-68;PMID: 10746014. Available from: <https://doi.org/10.1021/cc9800071>.
 39. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J Med Chem*. 2002 Jun 1;45(12):2615-23;PMID: 12036371. Available from: <https://doi.org/10.1021/jm020017n>.
 40. Stanzione F, Giangreco I, Cole JC. Use of molecular docking computational tools in drug discovery. In: *Progress in Medicinal Chemistry* [Internet]. Elsevier; 2021 [cited 2023 Sep 21]. p. 273-343;PMID: 34147204. Available from: <https://doi.org/10.1016/bs.pmch.2021.01.004>.
 41. Gu Z, Lin S, Yan W, Chen D, Zeng Z, Chen L, et al. Enhanced Water Solubility and Anti-Tumor Activity of Oleanolic Acid through Chemical Structure Modification. *Int J Mol Sci*. 2022 Oct 31;23(21):13291;PMID: 36362079. Available from: <https://doi.org/10.3390/ijms232113291>.
 42. Bank RPD. RCSB PDB;
 43. Liang D, Hao ZY, Zhang GJ, Zhang QJ, Chen RY, Yu DQ. Cytotoxic Triterpenoid Saponins from *Lysimachia clethroides*. *J Nat Prod*. 2011 Oct 28;74(10):2128-36;PMID: 21928797. Available from: <https://doi.org/10.1021/np2004038>.
 44. Li DQ, Wu J, Liu LY, Wu YY, Li LZ, Huang XX, et al. Cytotoxic triterpenoid glycosides (saikosaponins) from the roots of *Bupleurum chinense*. *Bioorg Med Chem Lett*. 2015 Sep;25(18):3887-92;PMID: 26259802. Available from: <https://doi.org/10.1016/j.bmcl.2015.07.053>.
 45. Wang O, Liu S, Zou J, Lu L, Chen L, Qiu S, et al. Anticancer Activity of 2 α , 3 α , 19 β , 23 β -Tetrahydroxyurs-12-en-28-oic Acid (THA), a Novel Triterpenoid Isolated from *Sinjackia sarcocarpa*. *Bauer JA*, editor. *PLoS ONE*. 2011 Jun 10;6(6):e21130;PMID: 21695177. Available from: <https://doi.org/10.1371/journal.pone.0021130>.
 46. Dang Thi TA, Kim Tuyet NT, Pham The C, Thanh Nguyen H, Ba Thi C, Doan Duy T, et al. Synthesis and cytotoxic evaluation of novel ester-triazole-linked triterpenoid-AZT conjugates. *Bioorg Med Chem Lett*. 2014 Nov;24(22):5190-4;PMID: 25442310. Available from: <https://doi.org/10.1016/j.bmcl.2014.09.079>.
 47. Yu L, Wang X, Wei X, Wang M, Chen L, Cao S, et al. Triterpenoid saponins from *Xanthoceras sorbifolia* Bunge and their inhibitory activity on human cancer cell lines. *Bioorg Med Chem Lett*. 2012 Aug;22(16):5232-8;PMID: 22801644. Available from: <https://doi.org/10.1016/j.bmcl.2012.06.061>.
 48. Quang TH, Ngan NTT, Minh CV, Kiem PV, Nhiem NX, Tai BH, et al. Anti-inflammatory Triterpenoid Saponins from the Stem Bark of *Kalopanax pictus*. *J Nat Prod*. 2011 Sep 23;74(9):1908-15;PMID: 21870831. Available from: <https://doi.org/10.1021/np200382s>.
 49. Gao J, Li X, Gu G, Liu S, Cui M, Lou HX. Facile synthesis of triterpenoid saponins bearing β -Glu/Gal-(1 \rightarrow 3)- β -GluA methyl ester and their cytotoxic activities. *Bioorg Med Chem Lett*. 2012 Apr;22(7):2396-400;PMID: 22406153. Available from: <https://doi.org/10.1016/j.bmcl.2012.02.032>.
 50. Mu LH, Huang CL, Zhou WB, Guo DH, Liu P. Methanolysis of triterpenoid saponin from *Ardisia gigantifolia* stapf. and structure-activity relationship study against cancer cells. *Bioorg Med Chem Lett*. 2013 Nov;23(22):6073-8;PMID: 24095094. Available from: <https://doi.org/10.1016/j.bmcl.2013.09.029>.
 51. Lu Y hong, Chen M cang, Liu F, Xu Z, Tian X ting, Xie Y, et al. Synthesis and Cytotoxic Activity of Novel C-23-Modified Asiatic Acid Derivatives. *Molecules*. 2020 Aug 14;25(16):3709;PMID: 32823913. Available from: <https://doi.org/10.3390/molecules25163709>.
 52. Zhang HJ, Zhang GR, Piao HR, Quan ZS. Synthesis and characterization of celastrol derivatives as potential anticancer

- agents. *J Enzyme Inhib Med Chem*. 2018 Jan 1;33(1):190-8;PMID: 29231066. Available from: <https://doi.org/10.1080/14756366.2017.1404590>.
43. Wang XY, Gao H, Zhang W, Li Y, Cheng G, Sun XL, et al. Bioactive oleanane-type saponins from the rhizomes of *Anemone taipaiensis*. *Bioorg Med Chem Lett*. 2013 Oct;23(20):5714-20;PMID: 23992864. Available from: <https://doi.org/10.1016/j.bmcl.2013.08.006>.
44. Du Z, Li G, Zhou X, Zhang J. Synthesis of MeON-Glycoside Derivatives of Oleanolic Acid by Neoglycosylation and Evaluation of Their Cytotoxicity against Selected Cancer Cell Lines. *Molecules*. 2021 Feb 2;26(3):772;PMID: 33540945. Available from: <https://doi.org/10.3390/molecules26030772>.
45. Moustafa GO, Shalaby A, Naglah AM, Mounier MM, El-Sayed H, Anwar MM, et al. Synthesis, Characterization, In Vitro Anticancer Potentiality, and Antimicrobial Activities of Novel Peptide-Glycyrrhetic-Acid-Based Derivatives. *Molecules*. 2021 Jul 28;26(15):4573;PMID: 34361728. Available from: <https://doi.org/10.3390/molecules26154573>.
46. Wu J, Zhang ZQ, Zhou XD, Yao QY, Chen ZL, Chu LL, et al. New Terpenoids from *Potentilla freyniana* Bornm. and Their Cytotoxic Activities. *Molecules*. 2022 Jun 7;27(12):3665;PMID: 35744788. Available from: <https://doi.org/10.3390/molecules27123665>.
47. Golbraikh A, Tropsha A. Beware of q2! *J Mol Graph Model*. 2002 Jan;20(4):269-76;PMID: 11858635. Available from: [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1).