

Quinazoline analogues as cytotoxic agents; QSAR, docking, and *in silico* studies

Leila Emami^{1,2}, Razieh Sabet^{1,*}, Soghra Khabnadideh^{1,2}, Zeinab Faghieh², and Parvin Thayori¹

¹Department of Medicinal Chemistry, Faculty of Pharmacy, Shiraz University of Medical Sciences, Shiraz, I.R. Iran.

²Department of Medicinal Chemistry, Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, I.R. Iran.

Abstract

Background and purpose: Synthesis and investigation of pharmacological activity of novel compounds are time and money-consuming. However, computational techniques, docking, and *in silico* studies have facilitated drug discovery research to design pharmacologically effective compounds.

Experimental approach: In this study, a series of quinazoline derivatives were applied to quantitative structure-activity relationship (QSAR) analysis. A collection of chemometric methods were conducted to provide relations between structural features and cytotoxic activity of a variety of quinazoline derivatives against breast cancer cell line. An *in silico*-screening was accomplished and new impressive lead compounds were designed to target the epidermal growth factor receptor (EGFR)-active site based on a new structural pattern. Molecular docking was performed to delve into the interactions, free binding energy, and molecular binding mode of the compounds against the EGFR target.

Findings/Results: A comparison between different methods significantly indicated that genetic algorithm-partial least-squares were selected as the best model for quinazoline derivatives. In the current study, constitutional, functional, chemical, resource description framework, 2D autocorrelation, and charge descriptors were considered as significant parameters for the prediction of anticancer activity of quinazoline derivatives. *In silico* screening was employed to discover new compounds with good potential as anticancer agents and suggested to be synthesized. Also, the binding energy of docking simulation showed desired correlation with QSAR and experimental data.

Conclusion and implications: The results showed good accordance between binding energy and QSAR results. Compounds Q₁-Q₃₀ are desired to be synthesized and applied to *in vitro* evaluation.

Keywords: Cytotoxic; Molecular docking; QSAR; Quinazoline.

INTRODUCTION

Among the nitrogen-containing compounds, the quinazoline ring was a very privileged and effective scaffold in pharmaceutical and medicinal chemistry; they have a broad spectrum of biological activities such as anticancer, diuretic, anti-inflammatory, anticonvulsant, antimicrobial, antiviral, antiparasitic, and antihypertensive effects (1). Also, among the different quinazoline scaffolds, 2-substituted-4(3H)-quinazolinone has been used as an attractive pharmacophore for drug design purposes (2,3). Quinazoline which is

substituted at C₄, C₆, and C₇ has been applied as one of the most significant classes of quinazoline-based epidermal growth factor receptor (EGFR) inhibitors (4). The quinazoline skeleton is also a substructure of natural purine bases, plant alkaloids, and several FDA-approved drugs such as prazosin, alfuzosin, erlotinib, gefitinib (Fig. 1) (5,6).

Access this article online



Website: <http://rps.mui.ac.ir>

DOI: 10.4103/1735-5362.323919

*Corresponding author: R. Sabet
Tel: +98-7132424127-8, Fax: +98-7132424126
Email: Sabet_r@sums.ac.ir

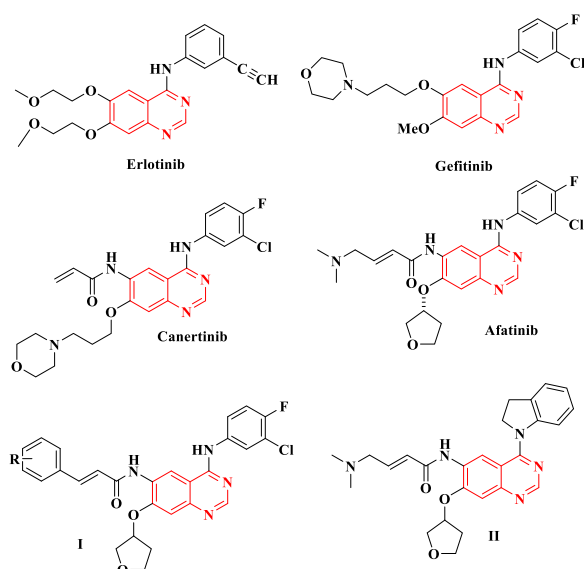


Fig 1. Previously reported epidermal growth factor receptor inhibitors bearing quinazoline scaffold.

Synthesis and investigation of pharmacological activity of novel compounds, usually take large amounts of expenditures and time. The use of computational techniques, docking, and *in silico* studies for designing pharmacologically effective compounds has opened a new approach to drug discovery research. Quantitative structure-activity relationships (QSAR) studies, as one of the important subjects in chemometrics, provide the ability to biological activities prediction for the novel or even non-synthesized compounds by medicinal chemists (7-11).

Linear QSAR models are mathematical equations that deliver good information for better explaining the mechanisms of action of the compounds, and proving the relationship between chemical structures and pharmacological activities.

The most important phase in constructing QSAR models is the suitable description of the structural and physicochemical properties of chemical structures (12-15). These properties called molecular descriptors have a high impact on the pharmacological properties of a compound (16-19). Molecular descriptors have been classified into different groups including physicochemical, constitutional, geometrical, topological, and quantum chemical descriptors. Dragon and Hyperchem as two famous computational packages are able to calculate more than 7000 of these parameters (20,21).

An important approach for the researchers is establishing a complete SAR of quinazoline derivatives of cytotoxic agents to modify the quinazoline moiety (22,23). In this paper, we investigated QSAR studies of some quinazoline derivatives which have been recently reported to exhibit cytotoxic activity against the MCF-7 cell line. Different QSAR models including multiple linear regressions were used to establish the relationship between descriptors and anti-breast cancer activity of compounds (16,24-26). Our QSAR models would be established mathematical equations between pharmacological activities and calculable parameters such as topological, quantum, physicochemical, stereochemical, or electronic parameters. In addition, molecular docking simulation was also done to reach the most favorable conformation and binding modes of all compounds as well as newly designed compounds towards EGFR as possible targets for their anticancer effect.

The molecular docking simulation help us to understand the possible interactions between the ligands and enzyme's active sites in detail and also helps to design novel potent inhibitors. In 2017, Tu *et al.* reported docking studies of new synthesized quinazoline bearing aryl semi carbazole (Fig. 1, structure I) and showed that tetrahydrofuran substituent was necessary for the anti-cancer activity of these compounds (27). The *in vitro* cytotoxic activities of new quinazoline compounds were determined against three human cancer cell lines (A549, MCF-7, and PC-3), and showing that the introduction of quinazoline derivatives bearing 2,3 dihydro-indole or 1,2,3,4-tetrahydroquinoline moiety (Fig. 1, structure II) could act appropriately to interact with the active site of the EGFR target (Fig. 1) (28).

MATERIAL AND METHODS

Data set and descriptor generation

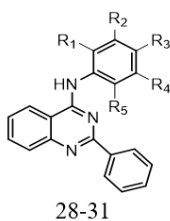
In this study, the experimental biological activities of anticancer activity against MCF-7 cell line were used (in terms of $-\log IC_{50}$), of a set of 81 quinazoline derivatives (29-33). The structural feature and antiproliferative values of these compounds were shown in Table 1.

Table 1. Chemical structure of quinazoline derivatives used in this study and their docking binding energy.

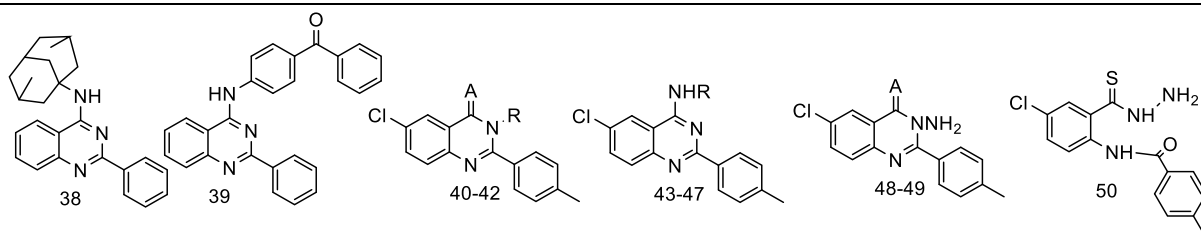
Order	R	pIC ₅₀	Binding energy (kcal/mol)
1-4			
5-8			
9-12			
13-14			
15-20			
1	NH ₂	1.6	-7.2
2	NHCOCH ₃	1.66	-8
3			
4		2.26	-8.8
5	H	1.87	-8.1
6	Cl	2.08	-8.7
7	F	2.16	-8.2
8	OCH ₃	1.7	-8
9	H	1.81	-8.9
10	Cl	1.54	-8.7
11	F	1.36	-8.5
12	OCH ₃	2.20	-8.8
13	H	1.79	-7.7
14	Cl	1.76	-8.5
15		1.61	-7.9
16		2.06	-8
17		2.18	-8.2
18		1.94	-8.9
19		2.11	-8.3
20		2.04	-8.4
21			
22-24			
25			
26			
27			

Table 1. (Continued)

Order	R	pIC ₅₀	Binding energy (kcal/mol)
21	-	1.76	-8.3
22	H	1.42	-8.7
23	Br	1.64	-8.5
24	F	1.54	-8.8
25	-	1.66	-8
26	-	1.99	-7.6
27	-	2.08	-8.4



Order	R ₁	R ₂	R ₃	R ₄	R ₅	pIC ₅₀	Binding energy (Kcal/mol)
28	H	CH ₃ -CH ₂ -	H	H	H	4.31	-9.6
29	H	H	H ₃ C-H ₂ C-O-	H	H	4.29	-8.7
30	H	H ₃ C-O-	H	H ₃ C-O-	H	4.64	-9.3
31	H	H ₃ C-O-	H ₃ C-O-	H ₃ C-O-	H	4.50	-8.7
32	CH ₃	H	H	H	NO ₂	4.33	-9.1
33	Cl	Cl	Cl	H	H	4.31	-9.9
34	Cl	H	Cl	Cl	H	4.28	-9.9
35	Cl	H	Cl	Cl	H	4.58	-9.5
36	H	H	Br	H	H	4.31	-9.6
37	Br	H	Br	H	H	4.37	-9.6



Order	R	A	pIC ₅₀	Binding energy (Kcal/mol)
38	-	-	4.44	-10.3
39	-	-	4.36	-9.9
40	O	H	4.73	-8.5
41	O	CH ₃	4.92	-8.6
42	S	H	4.93	-7.9
43	-	H	4.85	-8
44	-	OH	4.62	-8.2
45	-	NH ₂	4.87	-8.2
46	-		4.36	-8.7
47	-		4.84	
48	O	-	4.72	-7.8
49	S	-	4.51	-7.9
50	-		5.04	-6.9

Table 1. (Continued)

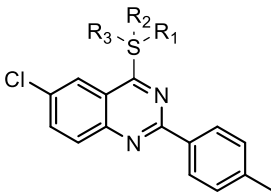
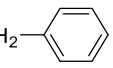
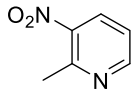
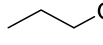
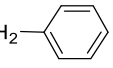



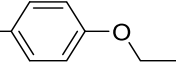
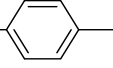
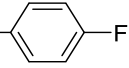
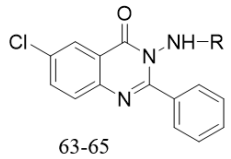
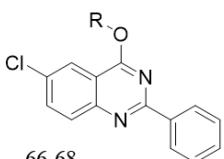
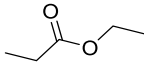
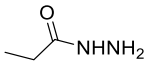
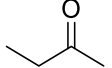
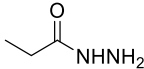
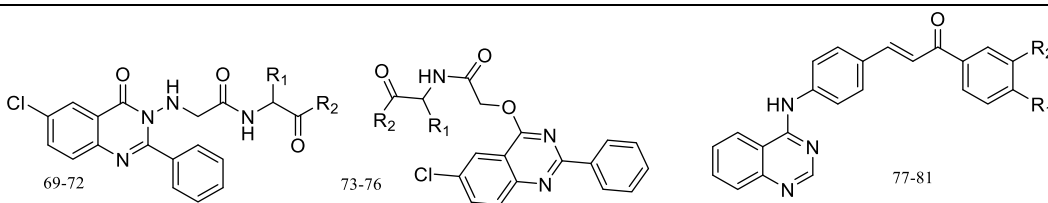
Order	R1	R2	R3	pIC ₅₀	Binding energy (Kcal/Mol)
 51-62					
51	—CH ₃	-	-	4.54	-7.9
52	—CH ₂ — 	-	-	4.59	-9.3
53	-H ₂ C-CN	-	-	4.98	-8.2
54	H	-	-	4.55	-7.9
55		-	-	5.06	-10.3
56		-	-	4.94	-8.0
57	—CH ₂ — 			5.01	-9.6
58	—CH ₂ —CO—NH ₂	-	-	4.78	-8.4
59	—CH ₂ —CO—NH— 	-	-	4.94	-8.1
60	—CH ₂ —CO—NH— 	-	-	4.46	-9.2
61	—CH ₂ —CO—NH— 	-	-	4.87	-10.0
62	—CH ₂ —CO—NH— 	-	-	4.47	-9.5
 63-65  66-68					
Order	R			pIC ₅₀	Binding energy (kcal/mol)
63	NH ₂			4.78	-8.1
64				4.93	-8.3
65				4.71	-9.3
66	H			4.79	-8.2
67				4.95	-8.3
68				4.47	-9

Table 1. (Continued)



Order	R1	R2	pIC ₅₀	Binding energy (Kcal/mol)
69	H	OCH ₃	4.91	-8.2
70	—H ₂ C—CH— 	OCH ₃	5.05	-8.2
71	H	—NHNH ₂	5.01	-8.9
72	—H ₂ C—CH— 	—NHNH ₂	5.12	-8.9
73	H	OCH ₃	4.99	-8.6
74	—H ₂ C—CH— 	OCH ₃	4.99	-8.6
75	H	—NHNH ₂	4.99	-8.8
76	—H ₂ C—CH— 	—NHNH ₂	5.05	-9.1
77	H	H	5.42	-9.8
78	Cl	H	5.69	-9.6
79	CF ₃	H	6.76	-10.9
80	—OCH ₃	—OCH ₃	-6.85	-10
81	—OCH ₃	H	6.79	-9.1

Hyperchem 8.0 software (Hypercube Inc. version 8.0.3) was used to draw the structures and optimization by semi-empirical AM1 calculations and Polak-Ribiere algorithm until the root mean square gradient of 0.01 kcal/mol (20). Molecular volume, molecular surface area, hydrophobicity (Log P), hydration energy, and molecular polarizability were achieved by Hyperchem software. The other descriptors such as topological, constitutional, and functional descriptors were achieved by the Dragon package (version 5.5), developed by the Milano chemometrics and QSAR group (21). The calculated descriptors are briefly introduced in Table 2. Finally, kenardston algorithm was used for the classification of the data set into calibration and test sets. This method is not accompanied by error and randomness.

Data screening and model building

Stepwise regression SPSS (version 22.0) software was used to analyze the experimental data to achieve effective descriptors. Then, the effective descriptors were hoarded in a data matrix that the number of molecules and descriptors were collected in rows and columns,

respectively. Multiple linear regressions (MLR), partial least squares (PLS), MLR with factor analysis (FA-MLR), and principal component regression analysis (PCRA) methods were utilized to derive the QSAR equations.

A cross-validation and external test set were used to evaluate the model's stability and prediction ability. For this purpose leave-one-out, leave-group-out were used (34). Also, the PLS regression with NIPALS-based algorithm existed in the chemometrics toolbox of MATLAB software (version 21 Math work Inc.) was used to obtain the optimum number of factors based on the Harland and Thomas F-ratio criterion (35).

Molecular modeling procedure

3D X-ray crystal structures of EGFR in complex with erlotinib (PDB ID: 1M17) were selected from Protein Data Bank (<http://www.rcsb.org>) based on the similarity of its co-crystal ligand to studied compounds (36,37). The structure of studied compounds was generated, minimized, and converted to pdbqt format. The pdbqt formats of the ligands were prepared by adding Gasteiger charges and

setting the degree of torsions by the AutoDock tools package (1.5.6). For the preparation of pdbqt format of protein, we removed cognate ligand, water molecules, added missing hydrogen atoms, and finally merged non-polar hydrogens according to their corresponding carbons (38,39). All preparation was performed by the AutoDock tools package (1.5.6). Docking procedure was done at the grid box

with a size of $70 \times 70 \times 70$ and the centre of $x = 20.14$, $y = 0.3$, $z = 52.2 \text{ \AA}$ by AutoDock Vina (1.1.2) (40). Using an in-house batch script (DOCKFACE) (41). The exhaustiveness was set to 100, and other docking parameters were set as default. The binding interactions of the docked compounds and the receptor were visualized by a fully automated protein-ligand interaction profiler (42).

Table 2. Brief description of some descriptors used in this study.

Descriptor type	Molecular description
Chemical	LogP (octanol-water partition coefficient), hydration energy (HE), polarizability (Pol), molar refractivity (MR), molecular volume (V), molecular surface area (SA).
Constitutional	Mean atomic van der Waals volume (MV), no. of atoms, no. of non-H atoms, no. of bonds, no. of heteroatoms, no. of multiple bonds (nBM), no. of aromatic bonds, no. of functional groups (hydroxyl, amine, aldehyde, carbonyl, nitro, nitroso, etc.), no. of rings, no. of circuits, no of H-bond donors, no of H-bond acceptors, no. of Nitrogen atoms (NN), chemical composition, sum of Kier-Hall electrotopological states (Ss), mean atomic polarizability (Mp), number of rotatable bonds (RBN), mean atomic Sanderson electronegativity (Me), number of Chlorine atoms (NCl), number of 9-membered rings (NR09), etc.
Topological	Molecular size index, molecular connectivity indices (X1A, X4A, X2v, X1Av, X2Av, X3Av, X4Av), information content index (IC), Sum of topological distances between F..F (T(F..F)), Ratio of multiple path count to path counts (PCR), Mean information content vertex degree magnitude (IVDM), Eigenvalue sum of Z weighted distance matrix (SEigZ), reciprocal hyper-detour index (Rww), Eigenvalue coefficient sum from adjacency matrix (VEA1), radial centric information index, 2D petijean shape index (PJI2), mean information index on atomic composition(AAC), Kier symmetry index(S0K), mean information content on the distance degree equality (IDDE), structural information content (neighborhood symmetry of 3-order) (SIC3), Randic-type eigenvector-based index from adjacency matrix (VRA1), sum of topological distances between N..N (T(N..N)), sum of topological distances between O..O(T(O..O)), etc.
Geometrical	3D-Balaban index (J3D), span R (SPAN), length-to-breadth ratio by WHIM (L/BW), sum of geometrical distances between N..N (G(N..N)), sum of geometrical distances between N..O (G(N..O)), sum of geometrical distances between O..O (G(O..O)), etc.
Mol-Walk	Molecular walk count of order 08 (MWC08), self-returning walk count of order 05 (SRW05), total walk count (TWC), etc.
Burden matrix	Highest eigenvalue n. 1 of burden matrix / weighted by atomic masses (BEHM1), highest eigenvalue n. 7 of Burden matrix / weighted by atomic masses (BEHM7), lowest eigenvalue n. 1 of Burden matrix / weighted by atomic masses (BELM1), highest eigenvalue n. 1 of Burden matrix / weighted by atomic van der Waals volumes (BELV1), highest eigenvalue n. 2 of Burden matrix / weighted by atomic Sanderson electronegativities (BEHE2), etc.
Galvez	Topological charge index of order 1 (GGI1), topological charge index of order 6 (GGI6), topological charge index of order 7 (GGI7), global topological charge index (JGT), etc.
2Dautocorrelation	Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic Sanderson electronegativities (ATS7E), Moran autocorrelation -lag 4 / weighted by atomic Sanderson electronegativities (MATS4E), Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic Sanderson electronegativities (ATS3E), Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic van der Waals volumes (ATS3V), etc.
Charge	Maximum positive charge (QPOS), partial charge weighted topological electronic charge (PCWTE), etc.
Aromaticity	Harmonic oscillator model of aromaticity index, RCI, Jug RC index aromaticity indices, HOMT; HOMA total (trial), etc.
Randic	DP0, molecular profile; SP0, shape profile; SHP, average shape profile index; etc.
RDF	Radial distribution function - 7.0 / unweighted (RDF070U), radial distribution function - 13.5 / unweighted (RDF135U), radial distribution function - 1.0 / weighted by atomic masses (RDF010M), radial distribution function - 3.0 / weighted by atomic masses (RDF030M), radial distribution function - 4.5 / weighted by atomic masses (RDF045M), radial distribution function - 12.5 / weighted by atomic masses (RDF125M), radial distribution function - 2.0 / weighted by atomic van der Waals volumes (RDF020V), radial distribution function - 8.5 / weighted by atomic van der Waals volumes (RDF085V), radial distribution function - 1.0 / weighted by atomic Sanderson electronegativities (RDF010E), etc.
3D-MoRSE	3D-MoRSE - signal 01 / unweighted (MOR01U) (01U,02U,...,32U), 3D-MoRSE - signal 01 / weighted by atomic van der Waals volumes (MOR01V) (01V,02V,...,32V), ect.

Table 2. (Continued)

Descriptor type	Molecular description
WHIM	First component symmetry directional WHIM index / weighted by atomic polarizabilities (G1P), second component symmetry directional WHIM index / weighted by atomic electrotopological states (G2S), D total accessibility index / weighted by atomic van der Waals volumes (DV), etc.
GETAWAY	H autocorrelation of lag 1 / lag2/ lag3 weighted by atomic Sanderson electronegativities (H1E,H2E,H3E), total information content on the leverage equality (ITH), R maximal autocorrelation of lag 3 / lag4 unweighted (R3U+,R4U+), R maximal autocorrelation of lag 6 / weighted by atomic masses (R6M+), R maximal autocorrelation of lag 5 / weighted by atomic van der Waals volumes (R5V+), R maximal autocorrelation of lag 1 / lag 4 weighted by atomic Sanderson electronegativities (R1E+), R maximal autocorrelation of lag 3 / weighted by atomic polarizabilities (R3P+), etc.
Functional	Number of total secondary C(sp3) (NCS), number of ring tertiary C(sp3) (NCRHR), number of secondary C(sp2) (n=CHR), number of tertiary amines (aliphatic) (NNR2), number of N hydrazines (aromatic) (nN-NPH), number of nitriles (aliphatic) (NCN), number of phenols (NOHPH), number of ethers (aromatic) (NRORPH), number of sulfures (NRSR), etc.
Atom-centred	CHR3 (C-003), CR4 (C-004), X--CR..X (C-034), Ar-C(=X)-R (C-039), R-C(=X)-X / R-C#X / X=C=X (C-040), X--CH..X (C-042), H attached to C1(sp3) / C0(sp2) (H-047), RCO-N</>N-X=X (N-072),R2S / RS-SR (S-107), etc.
Connectivity indices	X0 (connectivity index chi-0), connectivity index chi-1 (x1), average connectivity index chi-0 (XOA)
Information indices	Uindex (Balaban U index), IC0 (information content index), TIC0 (total information content index)
Edge adjacency indices	EEig01x (eigen value 01), EEig01r (eigen value 01 from edge)
Eigenvalue-based indices	Eig1v (leading eigenvalue from van der Waals weighted distance matrix), SEigm: eigenvalue sum from mass-weighted distance matrix eigen value-based indices

RESULTS

MLR analysis

The resulted QSAR models from different types of descriptors for the compounds (65 molecules as calibration and 16 molecules as prediction sets) are listed in Table 3. The stepwise selection was used to achieve MLR analyses.

Table 4 includes the statistical parameters of prediction that indicate the suitability of the proposed QSAR model based on MLR analysis of molecular parameters. The correlation coefficient of prediction is 0.97, which means that the resulted QSAR model could predict 97% of variances in the anti-breast cancer activity data. It has a root mean square error (RMSE) of 0.21.

Genetic algorithm-PLS analysis

Genetic algorithm (GA) method was used to select the variable to find more effective descriptors to improve the performance of PLS analysis. The data set (n = 81) was disorted in calibration (n = 65) and prediction set (n = 16) groups. A cross-validation procedure was used to achieve the optimum number of latent variables for each PLS model. In this work,

many different GA-PLS runs were conducted using the different initial set of populations (50-250), therefore, the most convenient GA-PLS model that resulted in the best fitness contained 15 descriptors including six constitutional descriptors (nI, nR10, nBnz, nArX, nO, nX), five functional descriptors (n ROR, nC=S, npyridine, nHAcc, nArNO₂), one chemical (SA), one resource description framework (RDF) parameter (RDF115v), one 2D autocorrelation parameter (MATS3m), and one charge (Qneg) parameter. The majority of these descriptors are constitutional indices. All of them being those obtained by different MLR-based QSAR models. The PLS estimate of the regression coefficients is shown in Fig. 2.

In order to investigate the relative importance of the variable that appeared in the final model, the variable important in projection (VIP) was employed (43). VIP values reflect the importance of terms in the PLS model. According to Erikson *et al.* X-variables (predictor variables) could be classified according to their relevance in explaining y (predicted variable), so that $VIP > 1.0$ and $VIP < 0.8$ mean highly or less influential, respectively, and $0.8 < VIP < 1.0$ means moderately influential (Fig. 3).

Table 3. The results of multiple linear regressions analysis with different types of descriptors.

Equations	Descriptors	Positive effect	Negative effect	R ²	F	Q ²	SE
Eq1	Constitutional	nF, nO, nCl, nI, nX, nAB, nR10, nR06		0.91	218.993	0.87	0.47
Eq2	Topological descriptors	T(O..O), PCR, Jhetv	T(O..I), JhetZ, SEigv, piID, T(F..I)	0.94	118.01	0.92	0.40
Eq3	BCUT descriptors	-	BEHm1	0.89	514.932	.088	0.52
Eq4	Galvz topol. charge in dices	JGI2, JGI5	JGI4, JGI6, GGGI3	0.72	31.46	0.68	0.86
Eq5	2D autocorrelations	MATS6m GATS4p ATS7m GATS1p MATS1m MATS1p	GATS1m ATS4e ATS3m ATS8e MATS3m	0.96	114.44	0.93	0.35
Eq6	Geometrical descriptors	G(N..O) G(N..S) G(N..F)	G(O..I) G(N..I) G(O..F)	0.91	100.78	0.73	0.49
Eq7	RDF descriptors	RDF115v RDF120m RDF020u	RDF020m RDF140v	0.94	161.59	0.92	0.39
Eq8	3D descriptors	MoRSE Mor11m Mor03m Mor28m Mor04v Mor15m Mor26v	Mor04m Mor13v Mor10m	0.92	72.61	0.88	0.47
Eq9	WHIM descriptors	E1s	E1m, Am	0.87	145.88	0.85	0.56
Eq10	GETAWAY descriptors	R3v+	R3p+ R8v+	0.88	156.77	0.4	0.55
Eq11	Fuctional counts	group n=CHR, nCb, nArX, nArNO2, nN-N, nROR, nSO2, nHAcc	nCONR2, nC=S, npyridine	0.82	54.33	0.77	0.69
Eq12	Atom-centred fragments	C-016 O-058	I-099 C-034 C-024 C-030	0.97	322.25	0.96	0.28

Table 4. Statistical parameters for testing prediction ability of the MLR, GA-PLS, PCR, and FA-MLR models

Model	R ²	R ² Loocv	RMSEcv	R ² p	RMSEp
MLR	0.97	0.9687	0.2809	0.97	0.21
GA-PLS	0.9803	0.9692	0.2788	0.9365	0.2125
PCR	0.93	0.919	0.4515	0.95	0.25
FA-MLR	0.97	0.9687	0.2809	1.00	-

MLR, Multiple linear regressions; GA-PLS, genetic algorithm-partial least squares; PCR, principal component regression; FA-MLR, MLR with factor analysis; R²: regression coefficient for calibration set; R²Loocv, regression coefficient for leave one out cross-validation; RMSEcv, root mean square error of cross-validation; R²p, regression coefficient for prediction set; RMSEp: root mean square error of prediction set.

FA-MLR and PCRA

In FA-MLR analysis, Factor Analysis was used to decrease the number of variables and also to identify the effective predictor variables and to decrease collinearities between them (44). PCRA was tried for the dataset along with FA-MLR (45). In this method, factor scores, as obtained from FA, were used as the predictor variables (44).

Based on the process explained in the experimental part, the following five-parametric equation is shown in Table 5.

$$Y = 4.752(\pm 0.170) + 3.237(\pm 0.122) nI + 1.814(\pm 0.294) \quad (1)$$

$$n = CHR + 0.134(\pm 0.030) nO - 0.069(\pm 0.17) nCONR_2 + 0.064(\pm 0.027) nN-N$$

$$R^2 = 0.97; S.E = 0.24; F = 25.18; Q^2 = 0.96; RMSEcv = 0.11$$

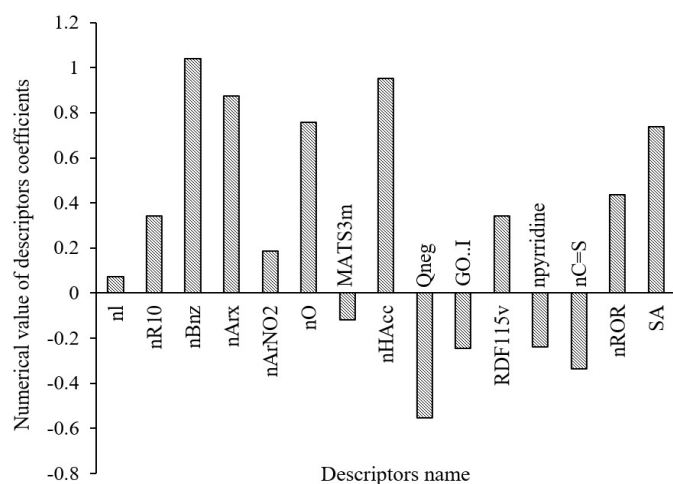


Fig. 2. Partial least squares regression coefficients for the variables used in the genetic algorithm-partial least squares model.

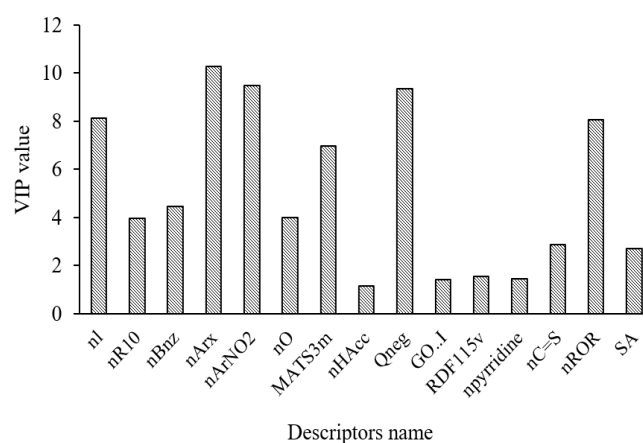


Fig. 3. Plot of VIP for the descriptors used in genetic algorithm-partial least squares model. VIP, variables important in projection.

Table 5. The results of multiple linear regressions analysis with factor analysis with different types of descriptors

Models	Unstandardized coefficients		Standardized coefficients	t	Sig.	R ²	F	Q ²	SE
	B	SE	Beta						
Constant	4.752	.170		27.892	.000	0.97	25.18	0.9687	0.24
nI	3.237	.122	.999	26.493	.000				
n=CHR	1.814	.249	.307	7.290	.000				
nO	.134	.030	.116	4.533	.000				
nCONR ₂	.069	.017	-.126	-4.021	.000				
nN-N	.064	.027	.080	2.342	.023				

Table 6. The results of principal component regression analysis.

	Unstandardized coefficients		Standardized coefficients	t	Sig.	R ²	F	Q ²	SE
	B	SE	Beta						
(Constant)	3.866	.054		70.939	.000	0.93	23.354	0.91	0.42
F1	-1.321	.051	-.887	-26.049	.000				
F4	.361	.049	.250	7.372	.000				
F2	-.420	.053	-.270	-7.940	.000				
F3	-.319	.055	-.197	-5.761	.000				

This equation could explain about 97% of the variance and predict 96% of the variance in pIC₅₀ data. It has an RSME of 0.15. This equation describes the effect of functional, constitutional and so on descriptors (nI, n = CHR, nO-O, nCONR₂ and nN-N) on anti-breast activity of the studied molecules. Multiple regression equation using stepwise selection method (PCRA), the following equation was gained (Table 6).

$$Y = 3.866 (\pm 0.054) - 1.321 (\pm 0.51) F1 + 0.361 (\pm 0.049) F4 - 0.420 (\pm 0.053) F2 - 0.319 (\pm 0.055) F3 \quad (2)$$

$$R^2 = 0.93; S.E. = 0.42; F = 23.35; Q^2 = 0.91; RMS_{cv} = 0.19$$

This equation could explain and predict 93% and 91% of the variances in pIC₅₀ data, respectively. The RSME of PCRA analysis was 0.19. PCRA equation is better than those derived from FA-MLR. While the data of this analysis showed acceptable prediction, we observed that the predicted values of some molecules are near to each other.

As it is shown in Table 5, in the case of each factor, the loading values for some descriptors are much higher than those of the others. These high values for each factor indicate that this factor delivers higher information about each index. It should be noted that all factors have

information from all descriptors but the contribution of the descriptor in different factors are not equal. For example, factors 1 and 2 have higher loadings for the chemical, atom-center, constitutional, functional, BCUT (modified burden eigenvalues) information, geometrical, 2D autocorrelations, and Walk and path counts indices whereas information about the weighted holistic invariant molecular (WHIM descriptor), connectivity indices, 3D MoRSE descriptors and functional descriptors are highly combined in factors 3 and 4.

To assess the robustness and applicability domain of all gained models, leverage as standard procedures was used (Table 7). Warning leverage (h^*) is another criterion to explain results. Leverage greater than warning leverage h^* means that the predicted response is the result of substantial extrapolation of the model and therefore may not be reliable (46).

Docking studies

Molecular docking simulation is the computational method of ligand binding to a receptor that indicates the affinity of the binding ligand to the active site of the target. Quinazoline derivatives are well-known as EGFR inhibitors (47). So, the docking study was done for EGFR kinase to predict the binding modes, affinities, and orientation of ligands at the active sites of the enzyme. In the

present research, molecular docking simulations were done on 81 molecules of data set as well as 30 designs of new compounds. The docking method was verified by docking of co-crystallized ligand (erlotinib) in EGFR active site and comparison of the results between docked and crystallized erlotinib conformations. RMSD of docking for erlotinib in comparison with its coordination in the crystal structure was 1.07. The free energy binding of docked compounds is summarized in Table 1.

Docking study on EGFR

All investigated complexes showed better docking binding energies than the reference drug (erlotinib) with a binding energy value of -5.8 kcal/mol which showed a good correlation with QSAR results, it means almost that the compounds with the lowest IC_{50} have the highest binding energy. The docked conformation of protein crystal structure (1M17) with erlotinib and three other quinazoline derivatives (33, 55, and 79), which possess strong binding energy, are shown in (Fig. 2). As depicted in Fig. 4, the nitrogen atom of the quinazoline moiety in compound 33 binds to the Thr 766 *via* hydrogen bond interactions, and some hydrophobic interactions with the residues Lys 721, Ala 719, Leu 764, Met 769, and Leu 820 were observed.

Table 7. Leverage (h) of the external test set molecules for different models. The last row (h^*) is the warning leverage.

Order	MLR	GA-PLS	PCR	FA-MLR
10	0.06361	0.140646	0.032966	0.087362
16	0.074537	0.347163	0.06082	0.140623
34	0.040178	0.152778	0.01619	0.055034
40	0.040848	0.295696	0.038652	0.017237
42	0.044381	0.370464	0.04964	0.018789
43	0.091655	0.781316	0.041538	0.044838
48	0.05129	0.294866	0.0391	0.02379
49	0.039529	0.386286	0.05183	0.017309
51	0.058363	0.598925	0.046114	0.042804
61	0.075355	0.238978	0.038749	0.051489
65	0.025045	0.241746	0.026166	0.036154
66	0.054479	0.154907	0.020783	0.037855
67	0.057342	0.51316	0.042428	0.047063
72	0.037762	0.350891	0.016021	0.032708
73	0.094482	0.391798	0.077436	0.134198
h^*	0.075483	0.191118	0.025299	0.123573

FA-MLR, Multiple linear regressions with factor analysis; GA-PLS, genetic algorithm-partial least squares; PCR, principal component regression.

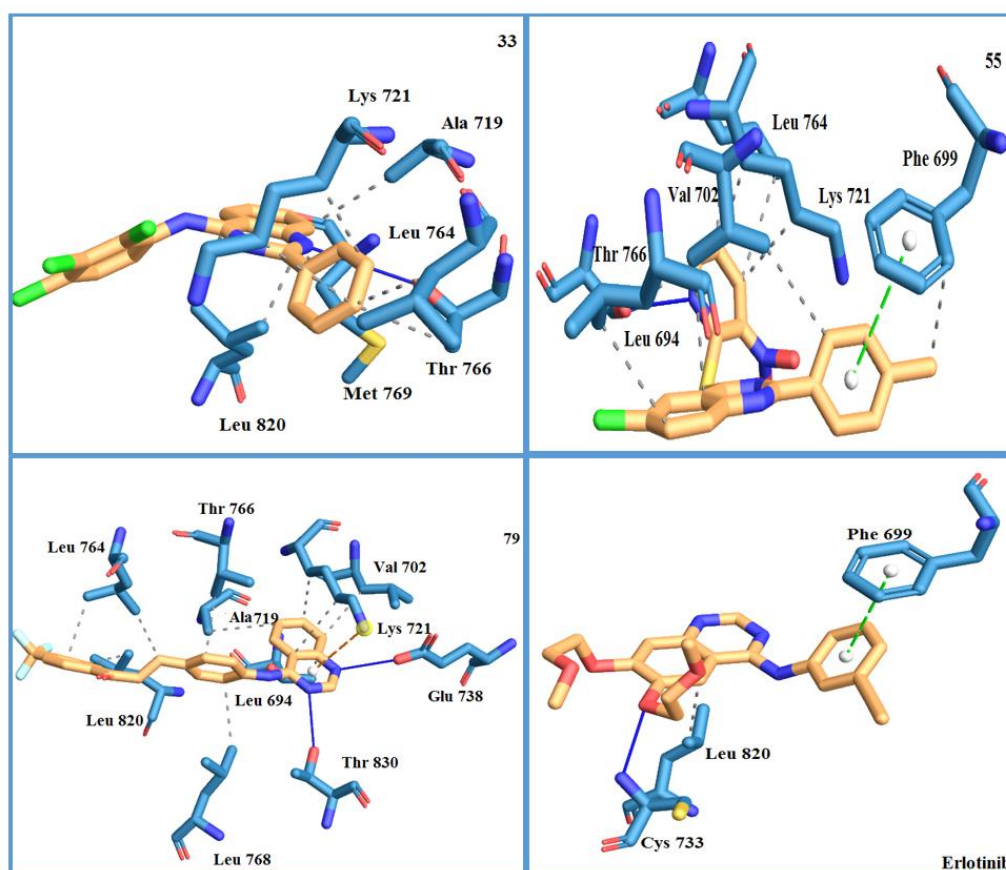


Fig. 4. The main interaction between the active site of epidermal growth factor receptor (PDB ID: 1M17) with compounds 33, 55, 79, and erlotinib.

In compound 55, the nitrogen atom formed hydrogen bond interaction with Thr 766 and the phenyl ring formed pi-stacking interaction with Phe 699 and also there is some hydrophobic interaction with Val 702, Leu 764, and Leu 694. Compound 79 can be summarized by hydrogen bond interactions with Glu 738 and Thr 830 and the pi-cation interaction between quinazoline ring and Lys 721. Also, interaction with Leu 699, 764, 766, 820, Val702, and Ala 719. To explain the binding mode of erlotinib, in brief, it should be announced that the oxygen atom of the erlotinib chain formed a hydrogen bond *via* Cys 733. In addition, pi-stacking interaction between the phenyl ring and Phe 699 was observed.

In silico screening

In silico methods can help to discover newly designed drugs before synthesized and offers a lot of benefits such as cost savings, time-consuming, and increase the speed of predict and achieve new pharmaceutical compounds. Virtual screening is a computational technique to search active compounds from very large

libraries by the effect of structural modification on binding energy and biological activity at the parent molecule. A new compound designed according to the best QSAR model and binding energy from docking simulation. We used the important descriptors selected in the GA-PLS model to design new active compounds because the GA-PLS model has the greatest statistical parameters. The *in silico* screen was applied by substituting different groups in different places. We also analyzed the robustness and applicability domain of the model by calculation of leverage value on the whole data set. The warning leverage (h^*) was found to be 0.66 and it is cleared that all of the compounds have a smaller leverage value (h) than warning leverage (h^*). The results of this investigation are summarized in Table 8. We designed 30 compounds (Q1-Q30) and biological activities based on the GA-PLS equation were achieved. And also, docking binding energy was obtained. From the results, the Q14 has the best activity and have a good candidate for anticancer agents.

Table 8. Structural modification of studied compounds and their predicted activities and docking binding energies for epidermal growth factor receptor inhibitory based on genetic algorithm-partial least squares equation.

Order	Structures	pIC ₅₀	leverage	BE	Order	Structure	pIC ₅₀	leverage	BE
Q1		5.42	0.4024	-9	Q16		5.39	0.1879	-10.3
Q2		5.53	0.3517	-9.4	Q17		5.46	0.2763	-10.4
Q3		5.14	0.2613	-10.1	Q18		5.35	0.2763	-10.2
Q4		5.04	0.2501	-9.5	Q19		5.29	0.4785	-10.2
Q5		5.56	0.3908	-9	Q20		5.23	0.4925	-10.1
Q6		5.32	0.1202	-10.1	Q21		5.08	0.4337	-10
Q7		5.45	0.2376		Q22		5.57		-10.3
Q8		5.16	0.4408	-9.3	Q23		5.04	0.415	-10.1
Q9		5.18	0.2422	-8.7	Q24		6.06	0.4467	-10.6
Q10		5	0.1207	-8.8	Q25		5.85	0.458	-10.3

Table 8. (Continued)

Order	Structures	pIC ₅₀	leverage	BE	Order	Structure	pIC ₅₀	leverage	BE
Q11		5.12	0.1676	-9	Q26		5.71	0.2092	-9.6
Q12		5.52	0.2379	-9.1	Q27		5.29	0.497	-9.6
Q13		5.33	0.2425	-8.9	Q28		5.57	0.0593	-9.9
Q14		7.02	0.2526	-10.7	Q29		5.34	0.4062	-9.8
Q15		5.27	0.0157	-8.5	Q30		5.08	0.1926	-9

BE, Binding energy.

Table 9. Physicochemical properties of some potent designed compounds.

Entry	MW	LogP	HBD	HBA	TPSA (Å ²)	n-RB	Lipinski violation
Q1	430.7	7.82	1	3	47.04	4	1
Q5	558.58	9.6	1	2	37.81	3	2
Q7	385.22	6.05	0	3	34.89	2	1
Q14	445.86	3.03	3	8	137.69	10	0
Q17	344.75	4.25	2	6	99.36	5	0
Q24	661.34	6.35	2	7	124.81	7	1
Q25	661.34	7.51	0	7	95.99	7	2
Q28	613.45	8.92	1	5	73.34	8	2
Rule of Lipinski	≤ 500	≤ 5	≤ 5	≤ 10	≤ 140	≤ 10	≤ 1

MW, Molecular weight; LogP, logarithm of partition coefficient between n-octanol and water; HBD, number of hydrogen bond donors; HBA, number of hydrogen bond acceptors; TPSA, topological polar surface area; nRB, number of rotatable bonds.

ADME and drug-likeness properties of designed compounds

The physicochemical properties of some potent designed compounds were calculated by <http://www.swissadme.ch/>, and the results are shown in Table 9. The molecular weight of compounds (Q7, Q14, and Q17) was in the accepted range. The log P of hit compound Q14 values is lower than 5 and displayed that it has desire lipophilicity. Furthermore, the hydrogen bond properties and total polar surface area, and rotatable bond number of compound Q14 is within

the acceptable limit. Generally, the data indicated that compound Q14 followed Lipinski's rule of five and Veber rule. The drug-likeness and absorption, distribution, metabolism, and excretion (ADME) properties of some potent compounds were derived from the preADMET online server (<http://preadmet.bmdrc.org/>). ADME properties including human intestinal absorption, *in vitro* caucasian colon adenocarcinoma cell permeability (Caco-2), skin permeability, *in vitro* plasma protein binding, and *in vivo*, blood-brain barrier penetration are shown in Table 10.

Table 10. *In silico* absorbance distribution metabolism and excretion of some potent designed compounds.

Entry	Absorption			Distribution	
	HIA (%)	<i>In vitro</i> Caco-2 cell permeability (nm/s)	<i>In vitro</i> skin permeability (log Kp, cm/h)	<i>In vitro</i> plasma protein bonding (%)	BBB (%)
Q ₁	97.04	53.01	-2.3	100	5.21
Q ₅	95.01	50.94	-2.3	100	17.15
Q ₇	99.17	49.64	-2.7	99.07	0.52
Q ₁₄	93.46	19.86	-4.09	80.06	0.01
Q ₁₇	95.37	19.39	-3.4	97.64	0.05
Q ₂₄	98.29	1.52	-2.5	94.6	1.39
Q ₂₅	98.69	20.73	-1.7	100	4.39
Q ₂₈	97.88	45.44	-1.8	92.56	0.34

HIA, Human intestinal absorption; BBB, *in vivo* blood-brain barrier penetration.

Human intestinal absorption analysis indicated that hit compound Q₁₄ indicated well human intestinal absorption, which causes quick absorption from the intestine to the bloodstream. The Caco-2 permeability parameters suggest that compound Q₁₄ has moderate permeability for penetration to biological membranes. Moreover, compound Q₁₄ binds moderately to plasma proteins, so it may be used as an oral dosage form. It has low blood-brain barrier penetration; so, it has a low neurotoxicity effect.

DISCUSSION

In MLR analysis the equation 1 (Table 2) explained the effect of constitutional descriptors on the cytotoxicity activity of these compounds. It showed that increasing the numbers of substitution halogen atoms (nF, nCl, nI) on the compounds results in an activity enhancement, such as molecular series 6-7, 9-14. It also showed that the iodine substitution on the 6th position of quinazoline ring would have better activity. It also explains the positive effect nR10 and nR06 (number of 6-membered rings) such as phenyl ring on activity (such as molecule series 19-24, 25-30, and 33-37 have good activity). This equation also showed the effect of O numbers on the activity. The Keto group on the quinazoline ring has better activity than thio keto group on the ring (molecules number 50 has low activity). It also explains the positive effect of nDB (number of double bonds) on the activity. Quinazoline and phenyl ring have multiple double bonds.

The result of the topological group counts parameter on cytotoxicity activity of the studied

molecules has been introduced by equation 2 of Table 2. It showed that among the topological descriptors T (O. O), PCR, and JhetV have positive effects and T (O..I), JhetZ, SEigv, piID, and T(F..I) have negative effects on cytotoxic activity of the compounds.

The equations 3-12 of Table 3 demonstrated the positive and negative effects of BCUT, Galvz topological, 2D autocorrelations, geometrical, RDF, 3D MoRSE, WHIM, GETAWAY, functional, and atom-centered fragments on the anti-breast cancer activity of these compounds. The MLR equation of Table 2 obtained from the pool of functional groups descriptors, E₁₁, explained the positive impact of the nCb (presence of benzene ring, especially with halogen atoms such as F and, Cl, is good for example compounds 33-37) n = CHR, nArX (number of X on the aromatic ring for example phenyl ring), nArNO₂ (number of nitro groups (aromatic)), nN-N(number of N hydrazines), nROR (number of ethers), nSO₂ (number of sulfone compounds 19-21), nHAcc(number of acceptor atoms for H-bonds (N, O, F)) on the anticancer activity. The nCONR2 (number of tertiary amides compound 2), nC=S (number of thioketo compound 42), npyridine (number of pyridines compound 21) had negative impacts on the anti-breast cancer activity. The negative sign of this group proposed that a decrease in the number of these descriptors resulted in an activity enhancement. This equation, which has a high statistical quality ($R^2 = 0.97$, $Q^2 = 0.96$).

The model extracted from PLS model has a high cross-validation statistic, and it also represents a high ability for modelling external test samples. It could explain and predict about 93% of variances in the cytotoxic activity of the

studied molecules. There is a close agreement between the experimental and predicted values of anticancer activity data to evaluate the significance of the 15 selected PLS descriptors in the protein tyrosine kinase inhibitory activity. The VIP analysis of PLS equation is shown in (Fig. 2). As it is observed, nArX, nI, Qneg, nROR, and nArNO₂ indices represent the most significant contribution in the resulted QSAR model. In addition, functional group parameters such as nO, n BnZ, and MATS3m are moderately influential parameters. The calculated leverage values of the prediction set

samples for all models and the warning leverage, as the threshold value for accepted prediction, are shown in Table 6. As seen, the leverages of all test samples were lower than h^* for all models. This means that all the predicted values were acceptable.

Finally, the FA-MLR and PCRA results shown in Table 11, demonstrate the four-factor loadings of the variables (after VARIMAX rotation) for the compounds tested for cytotoxic activity. As it is observed, about 62% of variances in the original data matrix could be explained by the selected four factors.

Table 11. Numerical values of factor loading numbers 1–4 for descriptors after varimax rotation

Descriptors	Components				Descriptors	Components			
	1	2	3	4		1	2	3	4
SA1	.086	.717	.458	-.045	G(N..I)	.714	.556	.214	-.016
nO	.263	.879	.058	-.074	G(O..F)	.136	-.052	.014	.677
nF	.052	-.030	.011	.729	G(O..I)	.722	.590	.157	-.048
nI	.878	.359	.216	-.034	RDF020u	-.028	.765	.240	.061
JhetZ	-.131	-.381	-.761	-.354	RDF020m	.870	.378	.232	-.027
Jhetv	-.395	-.576	-.600	-.132	RDF080m	.361	.672	.376	-.118
SEigv	-.180	-.886	-.051	-.180	RDF120m	.119	.189	.143	.310
piID	.167	.334	.673	.307	RDF115v	-.066	.499	.442	.255
PCR	.019	.107	.647	.576	RDF140v	-.081	.437	.154	-.059
T(O..O)	.243	.887	-.026	-.030	Mor03m	-.759	-.184	-.291	.134
T(O..I)	.678	.621	.144	-.041	Mor04m	.662	.248	-.124	.057
T(F..I)	.281	-.078	.012	.245	Mor10m	.452	.538	.293	.077
MWC04	.239	.610	.670	.007	Mor11m	-.780	-.313	-.210	.060
SRW05	.419	.212	.298	-.088	Mor15m	-.130	.129	.586	-.021
BEHm1	.893	.324	.236	-.043	Mor28m	-.634	-.053	-.099	.161
GGI3	.199	.611	.346	-.151	Mor04v	-.054	.193	-.512	-.012
JGI2	-.116	-.009	-.721	.198	Mor13v	-.364	-.394	-.555	-.307
JGI4	.260	.123	-.594	-.412	Mor26v	-.092	-.280	-.518	-.123
JGI5	-.142	.170	-.231	-.391	E1m	.906	.190	.108	.089
JGI6	.612	.212	.168	-.317	E1s	-.056	.218	-.173	.733
ATS3m	.830	-.042	-.116	-.014	Am	.335	.636	.476	.098
ATS7m	.845	.075	.227	-.028	R8v	.468	.245	.647	-.186
ATS4e	.451	.450	-.276	.080	R3v+	.750	-.084	.102	.161
ATS8e	.471	.615	.074	-.026	R3p	.479	.316	.554	.118
MATS1m	.733	-.165	-.223	.161	nCp	-.668	.075	-.090	-.312
MATS3m	-.339	-.169	-.052	.342	nCrH2	.102	.104	.167	-.206
MATS6m	.278	.263	-.227	.462	n=CHR	-.282	-.001	.227	.755
MATS1p	-.091	-.333	-.284	-.001	nCONR2	.636	.399	.035	-.096
GATS1m	.854	.272	.216	-.122	nROR	.328	.405	.357	-.072
GATS1p	-.866	.263	.017	-.018	C-016	-.222	.013	.351	.694
GATS4p	-.549	-.243	-.008	.311	C-024	.157	.092	.866	.147
Qneg	.107	.861	.269	.003	C-030	-.251	.067	.052	.725
SHP2	-.136	-.562	-.724	-.074	C-034	.161	.045	.425	-.108
G(N..O)	.033	.941	.034	.060	O-058	.217	.861	-.083	-.027
G(N..S)	.081	.554	-.056	-.002	I-099	.878	.359	.216	-.034
G(N..F)	.007	-.009	.001	.743	%Variance	22.449	18.603	12.759	8.368

The results of the docking studies indicated that the important amino acid in the active site of EGFR enzyme is Thr 766, Val 719 and 702. The majority interaction of studied compounds are hydrogen bond between substituent at quinazoline moiety with Thr 766 and the pi-hydrogen bond between receptor and phenyl group of ligands. Molecular docking simulations were also applied to the newly designed compounds to obtain binding mode and energy with EGFR target. The physicochemical and ADME properties of the designed compounds were also predicted. It was indicated that these compounds were compatible with the Lipinski and Veber rules and have desired pharmacokinetic characteristics. In future studies, 30 newly designed quinazoline compounds will be synthesized and subjected to cytotoxic laboratory evaluation.

CONCLUSION

The quantitative relationship between molecular descriptor and cytotoxic activity of different quinazoline derivatives was discovered by different chemometrics methods, such as MLR, FA-MLR, PCR, and GA-PLS. The reliability, accuracy, and predictability of the proposed models were evaluated by RSME of cross-validation, cross-validation, and the RMSE of prediction. A static comparison between different QSAR methods indicated that GA-PLS represented a good relationship between the chemical structure of quinazoline and anticancer activity for MCF-7 cell line. This study suggested that constitutional descriptors, functional descriptors, chemical, RDF parameter, 2D autocorrelation parameter, and charge parameter descriptors of molecules were important for the prediction of anticancer activity of quinazoline derivatives. *In silico* screening was employed to discover new compounds such as **Q14**, with good potential inhibitory as anticancer agents and suggested to be synthesized, furthermore, ADME and drug-likeness properties of designed compounds indicated that compound **Q14** is compatible with Lipinski and Veber rules and has desired pharmacokinetic properties. Also, the binding

energy of docking simulation showed a desire correlation with QSAR and experimental data.

Acknowledgments

This work was financially supported by Shiraz University of Medical Sciences, Shiraz, I.R. Iran, under Grant No. 97-01-36-16805.

Conflict of interest statement

The authors declared no conflict of interest in this study.

Authors' contribution

L. Emami and S. Khabnadideh analyzed the data and co-wrote the paper. R. Sabet designed the experiment, analyzed the data, and co-wrote the paper. Z. Faghieh and P. Thayori analyzed the data

REFERENCES

- Haghighijoo Z, Rezaei Z, Taheri S, Jani M, Khabnadideh S. A rapid and convenient method for synthesis of anilinoquinazoline: an improved synthesis of erlotinib derivatives. *Trends Pharmacol Sci.* 2015;1(3):173-178.
- Rahmannejadi N, Khabnadideh S, Yavari I. Synthesis, docking, and cytotoxic activities of novel 2-aryl-4-(arylamino) quinazolines. *Monatsh Chem.* 2018;149(11):2085-2092. DOI: 10.1007/s00706-018-2270-3.
- Rahmannejadi N, Zomorodian K, Faghieh Z, Faghieh Z, Khabnadideh S, Yavari I. Preparation, docking, antimicrobial and cytotoxic activities of 2-arylquinazolinones. *Br J Pharm Res.* 2017;20(3):1-11. DOI: 10.9734/JPRI/2017/38244.
- Emami L, Faghieh Z, Khabnadideh S, Rezaei Z, Sabet R, Harigh E, *et al.* 2-(Chloromethyl)-3-phenylquinazolin-4(3H)-ones as potent anticancer agents; cytotoxicity, molecular docking and *in silico* studies. *J Iran Chem Soc.* 2021;18:1877-1889. DOI: 10.1007/s13738-021-02168-1.
- Divar M, Zomorodian K, Bastan S, Yazdanpanah S, Khabnadideh S. Synthesis of some quinazolinone derivatives using magnetic nanoparticles-supported tungstic acid as antimicrobial agents. *J Iran Chem Soc.* 2018;15(7):1457-1466. DOI: 10.1007/s13738-018-1337-8.
- Faghieh Z, Rahmannejadi N, Sabet R, Zomorodian K, Asad M, Khabnadideh S. Synthesis of some novel dibromo-2-arylquinazolinone derivatives as cytotoxic agents. *Res Pharm Sci.* 2019;14(2):115-121. DOI: 10.4103/1735-5362.253358.

7. Ferreira MMC. Multivariate QSAR. In: Wang Z, editor. Encyclopedia of Physical Organic Chemistry. John Wiley & Sons; Vol. 3. 2016. pp. 1-38. DOI: 10.1002/9781118468586.epoc3013.
8. Nasab RR, Mansourian M, Hassanzadeh F. Synthesis, antimicrobial evaluation and docking studies of some novel quinazolinone Schiff base derivatives. Res Pharm Sci. 2018;13(3):213-220. DOI: 10.4103/1735-5362.228942.
9. Sabet R, Fassihi A, Hemmateenejad B, Saghaei L, Miri R, Gholami M. Computer-aided design of novel antibacterial 3-hydroxypyridine-4-ones: application of QSAR methods based on the MOLMAP approach. J Comput Aided Mol Des. 2012;26(3):349-361. DOI: 10.1007/s10822-012-9561-2.
10. Sabet R, Fassihi A, Moeinifard B. QSAR study of PEST derivatives as potent HIV-1 reverse transcriptase inhibitors. J Mol Graph Model. 2009;28(2):146-155. DOI: 10.1016/j.jmgm.2009.05.002.
11. Sadeghian-Rizi S, Sakhteman A, Hassanzadeh F. A quantitative structure-activity relationship (QSAR) study of some diaryl urea derivatives of B-RAF inhibitors. Res Pharm Sci. 2016;11(6):445-453. DOI: 10.4103/1735-5362.194869.
12. Jain SK, Yadav AK, Nayak P. 2D QSAR analysis on oxadiazole derivatives as anticancer agents. Int J Curr Pharm Res. 2011;3(4):27-33.
13. Sainy J, Sharma R. QSAR analysis of thiolactone derivatives using HQSAR, CoMFA and CoMSIA. SAR QSAR Environ Res. 2015;26(10):873-892. DOI: 10.1080/1062936X.2015.1095238.
14. Todeschini R, Consonni V. Handbook of Molecular Descriptors. Weinheim: WileyVCH; 2008. pp. 516-520. DOI: 10.1002/9783527613106.
15. Wan J, Zhang L, Yang G, Zhan CG. Quantitative structure-activity relationship for cyclic imide derivatives of protoporphyrinogen oxidase inhibitors: a study of quantum chemical descriptors from density functional theory. J Chem Inf Comput Sci. 2004;44(6):2099-2105. DOI: 10.1021/ci049793p.
16. Hamann M, Alonso D, Martín-Aparicio E, Fuertes A, Pérez-Puerto MJ, Castro A, et al. Glycogen synthase kinase-3 (GSK-3) inhibitory activity and structure-activity relationship (SAR) studies of the manzamine alkaloids. Potential for Alzheimer's disease. J Nat Prod. 2007;70(9):1397-1405. DOI: 10.1021/np060092r.
17. Akhtar MJ, Siddiqui AA, Khan AA, Ali Z, Dewangan RP, Pasha S, et al. Design, synthesis, docking and QSAR study of substituted benzimidazole linked oxadiazole as cytotoxic agents, EGFR and erbB2 receptor inhibitors. Eur J Med Chem. 2017;126:853-869. DOI: 10.1016/j.ejmech.2016.12.014.
18. Ray S, De K, Sengupta C, Roy K. QSAR study of lipid peroxidation-inhibition potential of some phenolic antioxidants. Indian J Biochem Biophys. 2008;45:198-205.
19. Hemmateenejad B, Emami L, Sharghi H. Multi-wavelength spectrophotometric determination of acidity constant of some newly synthesized Schiff bases and their QSPR study. Spectrochim Acta A Mol Biomol Spectrosc. 2010;75(1):340-346. DOI: 10.1016/j.saa.2009.10.037.
20. Mojaddami A, Sakhteman A, Fereidoonhezad M, Faghieh Z, Najdian A, Khabnadideh S, et al. Binding mode of triazole derivatives as aromatase inhibitors based on docking, protein ligand interaction fingerprinting, and molecular dynamics simulation studies. Res Pharm Sci. 2017;12(1):21-30. DOI: 10.4103/1735-5362.199043.
21. Todeschini R. Milano Chemometrics and QSAR Research Group. Available from: <https://michem.unimib.it/>.
22. Emami L, Faghieh Z, Sakhteman A, Rezaei Z, Faghieh Z, Salehi F, et al. Design, synthesis, molecular simulation, and biological activities of novel quinazolinone-pyrimidine hybrid derivatives as dipeptidyl peptidase-4 inhibitors and anticancer agents. New J Chem. 2020;44(45):19515-19531. DOI: 10.1039/D0NJ03774E.
23. Ashok UP, Kollur SP, Arun BP, Sanjay C, Suresh KS, Anil N, et al. *In vitro* anticancer activity of 4(3H)-quinazolinone derived Schiff base and its Cu (II), Zn (II) and Cd (II) complexes: preparation, X-ray structural, spectral characterization and theoretical investigations. Inorganica Chim Acta. 2020;511:119846. DOI: 10.1016/j.ica.2020.119846.
24. Fassihi A, Abedi D, Saghaei L, Sabet R, Fazeli H, Bostaki G, et al. Synthesis, antimicrobial evaluation and QSAR study of some 3-hydroxypyridine-4-one and 3-hydroxypyran-4-one derivatives. Eur J Med Chem. 2009;44(5):2145-2157. DOI: 10.1016/j.ejmech.2008.10.022.
25. Fassihi A, Sabet R. QSAR study of p56lck protein tyrosine kinase inhibitory activity of flavonoid derivatives using MLR and GA-PLS. Int J Mol Sci. 2008;9(9):1876-1892. DOI: 10.3390/ijms9091876.
26. Sabet R, Fassihi A. QSAR study of antimicrobial 3-hydroxypyridine-4-one and 3-hydroxypyran-4-one derivatives using different chemometric tools. Int J Mol Sci. 2008;9(12):2407-2423. DOI: 10.3390/ijms9122407.
27. Tu Y, OuYang Y, Xu S, Zhu Y, Li G, Sun C, et al. Design, synthesis, and docking studies of afatinib analogs bearing cinnamamide moiety as potent EGFR inhibitors. Bioorg Med Chem. 2016;24(7):1495-1503. DOI: 10.1016/j.bmc.2016.02.017.
28. OuYang Y, Zou W, Peng L, Yang Z, Tang Q, Chen M, et al. Design, synthesis, antiproliferative activity and docking studies of quinazoline derivatives bearing 2,3-dihydro-indole or 1,2,3,4-tetrahydroquinoline as potential EGFR inhibitors. Eur J Med Chem. 2018;154:29-43. DOI: 10.1016/j.ejmech.2018.05.006.

29. Abbas SE, Barsoum FF, Georgey HH, Mohammed ER. Synthesis and antitumor activity of certain 2,3,6-trisubstituted quinazolin-4 (3*H*)-one derivatives. *Bull Fac Pharm Cairo Univ.* 2013;51(2):273-282. DOI: 10.1016/j.bfopcu.2013.08.003.
30. Abdel-Fattah HA, El-Etrawy AS, Gabr NRM. Synthesis and biological evaluation of some new 1,3,4-oxa, thiadiazole and 1,2,4-triazole derivatives attached to benzimidazole. *Int J Pharm Chem.* 2014;04(03):112-118. DOI:10.7439/IJPC.V4I3.91.
31. El-Azab AS, Al-Omar MA, Alaa AM, Abdel-Aziz NI, Magda AA, Aleisa AM, *et al.* Design, synthesis and biological evaluation of novel quinazoline derivatives as potential antitumor agents: molecular docking study. *Eur J Med Chem.* 2010;45(9):4188-4198. DOI: 10.1016/j.ejmech.2010.06.013.
32. Ghorab MM, Alsaid MS. Novel 4-aminoquinazoline derivatives as new leads for anticancer drug discovery. *Acta Pharm.* 2015;65(3):299-309. DOI: 10.1515/acph-2015-0021.
33. Madhavi S, Sreenivasulu R, Yazala JP, Raju RR. Synthesis of chalcone incorporated quinazoline derivatives as anticancer agents. *Saudi Pharm J.* 2017;25(2):275-279. DOI: 10.1016/j.jsps.2016.06.005.
34. Koljonen J, Nordling TE, Alander JT. A review of genetic algorithms in near infrared spectroscopy and chemometrics: past and future. *J Near Infrared Spec.* 2008;16(3):189-197. DOI: 10.1255/jnirs.778.
35. Sabet R, Fassihi A, Saghaie L. Octanol-water partition coefficients determination and QSPR study of some 3-hydroxy pyridine-4-one derivatives. *J Pharm Res Int.* 2018;22(4):1-15. DOI: 10.9734/JPRI/2018/41142.
36. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* 2015;43(Web Server issue):W443-W447. DOI: 10.1093/nar/gkv315.
37. Fereidoonnehad M, Faghih Z, Mojaddami A, Rezaei Z, Sakhteman A. A comparative QSAR analysis, molecular docking and PLIF studies of some *N*-arylphenyl-2,2-dichloroacetamide analogues as anticancer agents. *Iran J Pharm Res.* 2017;16(3):981-998. PMID: 29535790.
38. Emami L, Sabet R, Sakhteman A, Zade MK. Quantitative structure-activity relationship and molecular docking studies of imidazolopyrimidine amides as potent dipeptidyl peptidase-4 (DPP4) inhibitors. *J Pharm Res Int.* 2019;27(6):1-15. DOI: 10.9734/jpri/2019/v27i630186.
39. Sabet R, Khabnadideh S, Fathi D, Emami L. QSAR and docking study of isatin analogues as cytotoxic agents. *J Pharm Res Int.* 2019;27(5):1-22. DOI: 10.9734/jpri/2019/v27i530184.
40. Morris GM, Huey R, Olson AJ. Using autodock for ligand-receptor docking. *Curr Protoc Bioinformatics.* 2008;8:Unit 8.14. DOI: 10.1002/0471250953.bi0814s24.
41. Zarenezhad E, Esmailzadeh S, Behrouz S, Emami L. Copper Schiff base complex as a new immobilized heterogeneous catalyst: experimental, theoretical, biological and docking study. *J Iran Chem Soc.* 2021;18:283-295. DOI: 10.1007/s13738-020-02024-8.
42. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules.* 2015;20(7):13384-13421. DOI: 10.3390/molecules200713384.
43. Lo YC, Rensi SE, Tornø W, Altman RB. Machine learning in cheminformatics and drug discovery. *Drug Discov Today.* 2018;23(8):1538-1546. DOI: 10.1016/j.drudis.2018.05.010.
44. van de Waterbeemd H, Timmerman H, Mannhold R, Krogsgaard-Larsen P. *Chemometric Methods in Molecular Design.* John Wiley & Sons; 2008. pp. 1-379. DOI: 10.1002/9783527615452.
45. Mahani NM, Zarandi AM, Horzadeh A. QSAR studies of novel iminochromene derivatives as carbonyl reductase 1 (CBR1) inhibitors. *Marmara Pharm J.* 2018;22(2):227-236. DOI: 10.12991/mpj.2018.60.
46. Qian Y, Wang Y, Sa R, Yan H, Pan X, Yang Y, *et al.* Metabolic fingerprinting of *Angelica sinensis* during growth using UPLC-TOFMS and chemometrics data analysis. *Chem Cent J.* 2013;7(1):42-51. DOI: 10.1186/1752-153X-7-42.
47. Pasha FA, Muddassar M, Srivastava AK, Cho SJ. *In silico* QSAR studies of anilinoquinolines as EGFR inhibitors. *J Mol Model.* 2010;16(2):263-277. DOI: 10.1007/s00894-009-0534-x.