University of Basra
Collage of pharmacy
Department of pharmaceutical chemistry

QSAR Study of 2-Substituted Aminopyridopyrimidin-7-one derivatives As Tyrosine kinase Inhibitors

Guided by:
Dr. Rita S. Elias
Leaqaa A. Raheem

Prepared by:
Mahmood Murtada Mohammed
Abbas Abdulamir Taban

Abstract

QSAR techniques increase the probability of success and reduce time and coast in drug discovery process. The study presented QSAR investigation of 2-Substituted Aminopyridopyrimidin-7-one derivatives As Tyrosine kinase Inhibitors. A substance that blocks the action of enzymes called tyrosine kinases. Tyrosine kinases are a part of many cell functions, including cell signaling, growth, and division. These enzymes may be too active or found at high levels in some types of cancer cells, and blocking them may help keep cancer cells from growing. Molecular descriptors that used in this study are molecular weight (Mw) , logarithm of partition coefficient (Clog P) , log S , energy of high occupied molecular orbital (εHOMO), energy of low unoccupied molecular orbital (εLUMO) and CMR. Several models for the prediction of biological activity have been drawn up by using the multiple regression technique. In this study several QSAR equations were formulated. R^2 values were in the range 0.87 – 0.91 while SE values were in the range0.10-0.12. One model with bi- parametric linear equation with R^2 value 0.87. Five models with tri- parametric linear equation with R^2 values of 0.879- 0.91 and 0.904 were presented. The biological activity of the studied compounds can be modeled with quantum chemical molecular descriptors.

Introduction:

QSAR (Quantitative Structure-Activity Relationships) is a mathematical relationship between a biological activity of a molecular system and its geometric and chemical characteristics.

QSAR attempts to find consistent relationship between biological activity and molecular properties, so that these “rules” can be used to evaluate the activity of new compounds [(1)].

QSAR regression models relate a set of "descriptors" variables (X) to the potency of the response variable (Y)

A QSAR has the form of a mathematical model:

Activity = ∫ (physicochemical properties and/or structural propertied)+ Error

The biological activity of molecules (therapeutic effect or toxic effect) is usually measured in assays to establish the level of inhibition of particular signal transduction or metabolic pathways. (The concentration of a substance required to give a certain biological response).

Physicochemical and structural properties are represented by molecular descriptors can be calculated by computational methods.

The error includes model error (bias) and observational variability, that is, the variability in observations even on a correct model.

Drug discovery often involves the use of QSAR to identify chemical structures that could have good inhibitory effects on specific targets and have low toxicity (non-specific activity) [2].

Basic Requirements in QSAR Studies [3]:

1- All analogues belong to a congeneric series.
2- All analogy exerts the same mechanism of action.
3- All analogues bind in a comparable manner.
4- The effects of isosteric replacement can be predicted.
5- Binding affinity is correlated to interaction energies.
6- Biological activities are correlated to binding affinity

Descriptors in QSAR Studies [4]:

Descriptors are information rich variables defining molecules.

Descriptors are grouped in two categories:

(a) Whole molecules descriptors

Whole molecule descriptors are indicator properties of the entire molecule itself (e.g., molecular weight, molar refractivity).

(b) Fragment descriptors

Fragment descriptors are calculated based on the constituent atoms or substituent functional groups of the molecule. Substituent constants like the hydrophobicity constant (π) and the Hammett constant (σ) fall in this category.

Advantages of QSAR [5]:

1- Reduce animal tests and the use of animals.
2- Cost saving potential.
3- Collect data in a short period of time.

Applications of QSAR [6]:

1- Provides an understanding of the effect of structure on activity.
2- There is also the potential to make predictions leading to the synthesis of novel analogues.

Tyrosine kinase Inhibitors

A **tyrosine kinase inhibitor** (**TKI**) is a pharmaceutical drug that inhibits tyrosine kinases. Tyrosine kinases are enzymes responsible for the activation of many proteins by signal transduction cascades. The proteins are activated by adding a phosphate group to the protein (phosphorylation), a step that TKIs inhibit. TKIs are typically used as anticancer drugs. For example, they have substantially improved outcomes in chronic myelogenous leukemia. They have also been used to treat other diseases, such as idiopathic pulmonary fibrosis.

They are also called **tyrphostins**, the short name for "tyrosine phosphorylation inhibitor", originally coined in a 1988 publication,[1] which was the first description of compounds inhibiting the catalytic activity of the epidermal growth factor receptor (EGFR).

Tyrosine kinase enzymes (TKs) can be categorized into receptor tyrosine kinases (RTKs), non-receptor tyrosine kinases (NRTKs), and a small group of dual-specificity kinases (DSK) which can phosphorylate serine, threonine, and tyrosine residues. RTKs are transmembrane receptor that includes vascular endothelial growth factor receptors (VEGFR), platelet-derived growth factor receptors (PDGFR), insulin receptor (InsR) family, and the ErbB receptor family, which includes epidermal growth factor receptors (EGFR) and the human epidermal growth factor receptor-2 (HER2) [(7)].

Mechanism of action of TKI

As a whole, tyrosine kinases phosphorylate specific amino acids on substrate enzymes, which subsequently alter signal transduction leading to downstream changes in cellular biology. The downstream signal transduction set off by TKs can modify cell growth, migration, differentiation, apoptosis, and death. Constitutive activation or inhibition, either by mutations or other means, can lead to dysregulated signal cascades, potentially resulting in malignancy and other pathologies. Therefore, blocking these initial signals via TKIs can prevent the aberrant action of the mutated or dysfunctional TKs.

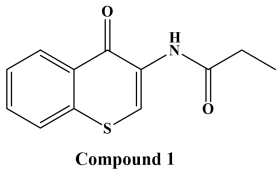
Despite the diverse primary amino acid sequences, human kinases share similar 3D structures, particularly when it comes to the ATP-binding pocket located in the catalytically active region. The starting amino acid sequence (ASP-Phe-Gly or) of the flexible activation loop that controls access to the activation site is also typically conserved.

Kinase inhibitors are either irreversible or reversible. The irreversible kinase inhibitors tend to covalently bind and block the ATP site resulting in irreversible inhibition. The reversible kinase inhibitors can further subdivide into four major subtypes based on the confirmation of the binding pocket as well as the .

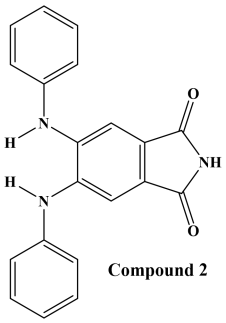
Pyrimidin derivatives

Epidermal growth factor receptors are a type of receptor tyrosine kinase.
a protein found on certain types of cells that binds to a substance called epidermal growth factor. The epidermal growth factor receptor protein is involved in cell signaling pathways that control cell division and survival. mutations (changes) in the EGFR gene cause epidermal growth factor receptor proteins to be made in higher than normal amounts on cancer cells. This causes cancer cells to divide more rapidly. EGFR is overexpressed in numerous tumors, including those derived from brain, lung, bladder colon, breast, head, and neck.an inhibitor that inhibits
EGFR TK has potential therapeutic value. Such agents have been extensively studied, especially in the pharmaceutical industry, as potential anticancer agents. Pyrimidin derivatives such as Aminopyridopyrimidin act as Inhibitors of Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase [7].

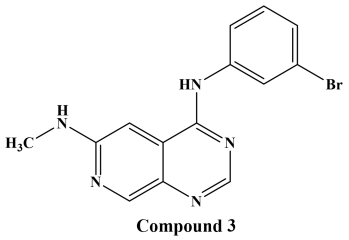
A number of inhibitors of protein TKs have been carried out toward suppression of the intracellular tyrosine phosphorylation of the EGFr TK. These compounds include highly potent and specific ATP-competitive inhibitors of the benzothiopyranones (e.g compound 1) and dianilinophthalimides (e.g. compound 2), and pyrido[3,4-d] pyrimidine (e.g. compound 3).



Compound 1



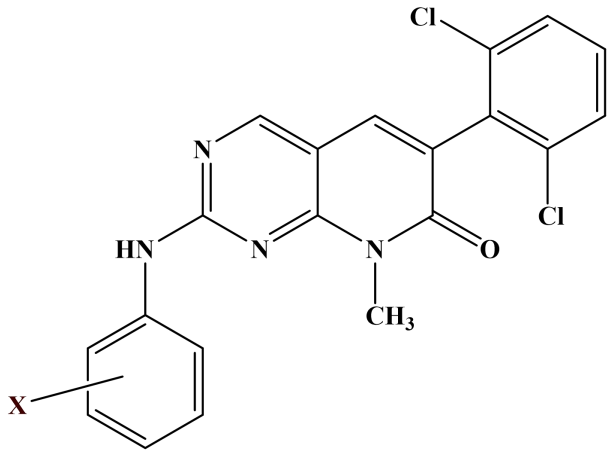
Compound 2



Compound 3

Compound 3 had an IC_{50} of 0.008 nM for inhibition of substrate phosphorylation, it was a competitive inhibitor of the EGFR with respect to ATP and inhibited EGFR autophosphorylation in A431 human epidermoid carcinoma cells with an IC_{50} of 13 nM. Compound 3 was also active against other members of the EGFR family, with IC50s of 49 and 52 nM respectively for inhibition of heregulin-stimulated

autophosphorylation in SK-BR-3 and MDA-MB-453 breast carcinomas. It also had a good activity against members of the EGFR family, while not influencing the function of related tyrosine kinases. The above properties of compound 3 make it an attractive lead compound for further development. A number of derivatives of compounds 3 had been synthesized and studied by **Klutchko** et al to find their inhibitory activities against these PTKs as shown in Figure1 [8].



X= H, 3-Br, 4-CH_3, 2-OCH_3, 3-OCH_3, 4-OCH_3, 3-OH, 3-CH_2OH, 3-CH_{3-4}-OCH_3, 3,5-di-OCH_3, 4-$(CH_2)_3CO_2C_2H_5$

Figure 1: The chemical structures of pyrido[3,4-d] pyrimidine derivatives

In this work, we have tried to investigate the relationship between structure and antitumor activity of a series of pyrido[3,4-d] pyrimidine derivatives (i.e. rationalize their biological activities using QSAR).

Materials and methods

The pyrido[3,4-d] pyrimidine derivatives have been taken with their reactivity from literature (**Klutchko** et al 1998).[8]

All calculations were done on personal laptop. All geometries of the pyrido[3,4-d] pyrimidine derivatives are minimized with the semi-empirical RM1 Hamiltonian. Molecular descriptors for the studied compounds, the hydrophobicity parameter of the whole molecule (Clog P), Mw, εHOMO and εLUMO were calculated using HyperChem8.5 program after molecular geometries were optimized first by the molecular mechanic method, and then by the semi- empirical RM1 Hamiltonian (Fig. 2).

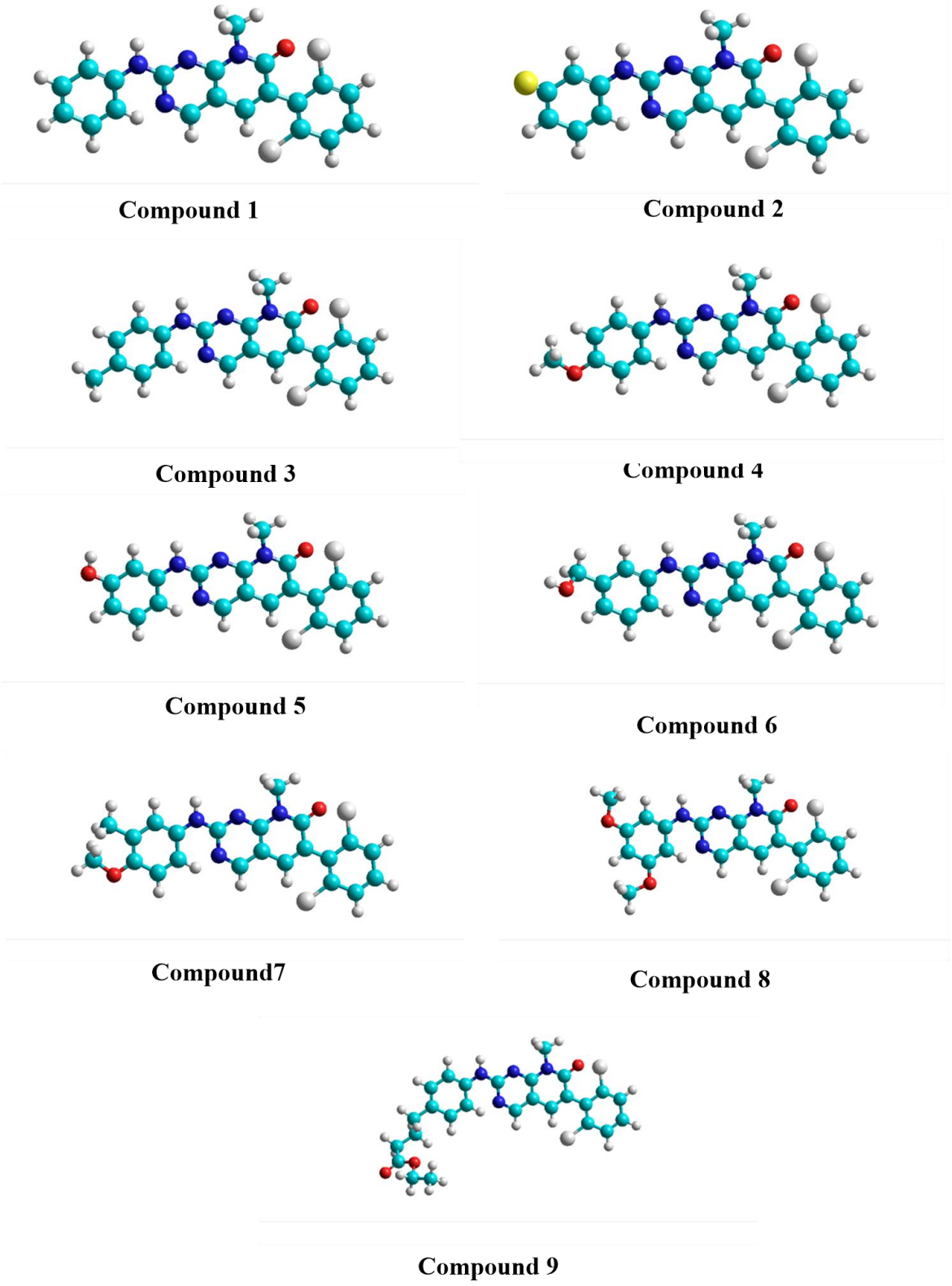


Figure 2: The semi- empirical RM1 Hamiltonian optimized structure of pyrido[3,4-d] pyrimidine derivatives

Multiple Linear Regression Analysis (MLR)

DATA Fit 9.1 program; have been used for multiple linear regression (MLR) analysis. The Quantum chemical descriptors used as independent variables and the biological activities values as the dependent variable. In the statistical analyses, the systematic search was performed to determine the significant descriptors.

Various regression equations have been developed by using selected quantum chemical descriptors. The best fitted regression equations have been used for the calculation of biological activities (BA pred).

In Model 1, the use of one descriptor (Clog P, M.W, HOMO, LUMO, ΔE and logS) as independent variable all gave poor correlation except Clog P give very nice correlation . In models 2-6, the use of two descriptors Clog p with ΔH, MW, and εLUMO, εHOMO, and log S and CMR as independent variables gave nice correlation.

RESULTS AND DISCUSSION

The general formulas of the chemical structures of pyrido[3,4-d] pyrimidine derivatives are shown in Figure 3. Chemical structures and experimental biological activities for studied compound are represented as IC_{50} (μM) are gathered in Table 1. The IC_{50} (μM) are within the range 0.03-0.7. The calculated molecular descriptors and biological activities are presented as log 1/C are gathered in table (2).

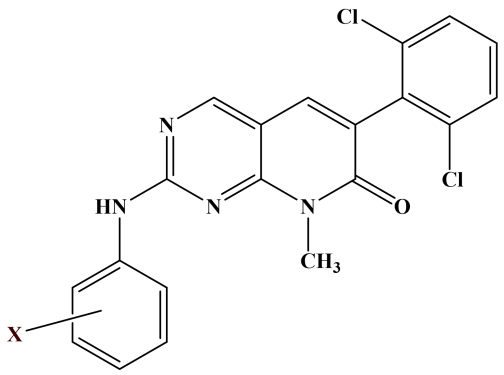


Figure 3: The general formula for chemical structures of the pyrido[3,4-d] pyrimidine derivatives

Table 1: The chemical structures of pyrido[3,4-d] pyrimidine derivatives and their observed activities as IC_{50} (μM) against Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase

No.	X	**EGFr** IC_{50} (μM)	Log 1/C (M)
1	H	0.26	6.59
2	3-Br	0.67	6.17
3	4-Me	0.48	6.32
4	2-OMe	0.66	6.18
5	3-OMe	0.03	7.52
6	4-OMe	0.22	6.66
7	3-OH	0.15	6.82
8	3-CH_2OH	0.09	7.05
9	3-Me,4-OMe	0.25	6.60
10	3,5-di-OMe	0.33	6.48
11	4-(CH2)3CO2C2H5	0.70	6.16

Table 2: Calculated molecular descriptors, observed activity against Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase

No.	X	Obsd.	CLog P	M.w	Log S	HOMO	LUMO	ΔE	CMR
1	H	6.59	4.79	397.26	-6.60	-8.53968	-1.13793	7.401748	10.7941
2	3-Br	6.17	5.85	476.15	-7.42	-8.68322	-1.26313	7.420094	11.5711
3	4-Me	6.32	5.48	411.28	-6.95	-8.46216	-1.10804	7.354119	11.2579
4	4-OMe	6.66	4.90	427.28	-6.65	-8.35873	-1.11585	7.242877	11.411
5	3-OH	6.82	4.31	413.26	-644	-8.37298	-1.32882	7.044159	10.9472
6	3-CH2OH	7.05	3.94	427.28	-6.13	-8.64181	-1.18286	7.458948	11.411
7	3-Me,4-OMe	6.60	5.40	441.31	-6.863	-8.150362	-1.22301	6.927347	11.8748
8	3,5-di-OMe	6.48	4.88	457.31	-6.72	-8.55362	-1.14796	7.405656	12.0279
9	4-(CH2)3CO2C2H5	6.16	6.40	511.40	-7.85	-8.53968	-1.13793	7.401748	13.9847

Log P; Partition coefficient; MW: Molecular weight (amu); ΔE: Energy gap (Kcal/mol)**;** εHOMO: Energy of HOMO(eV) ; εLUMO: energy of LUMO(ev); Log 1/c: observed activity expressed by log $1/IC_{50}$;

Klutchko et al. reported the inhibitory activity of analogues of 2-(X-phenylamino)-6-(2,6-di-Cl-phenyl)-8-Me-8*H*-pyrido[2,3-*d*] pyrimidin-7-ones. The various derivatives were obtained by different X-substituents (Table 1). They were tested for their ability to inhibit phosphorylation of a synthetic glutamate-tyrosine polymer by recombinant (human) EGFR-TK. we tried to correlate the activity of the compounds 1-9 represented by $\log(1/C_{50})$ as Inhibitors of Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase with the molecular descriptors, ClogP, M.W, HOMO, LUMO, ΔE and logS, CMR).

Six models were predicted in this study and been building up with the use of the following descriptors: Clog P, molecular weight (Mw), hydration energy log S, (ΔE) energy gap(eV) and energy of HOMO (ε_{HOMO}) (eV), and LUMO (ε_{LUMO}) (eV) and CMR.

The first model is bi-parametric equation (Eq.1) with one descriptor involve: Clog P, and the predicted values are reported under Table 4.

log1/C = -0.36 (±0.12)Clog P + 8.38(± 0.63).....(1)

n = 9, R^2= 0.87 , SE = 0.11 , F = 48.76

In this models' negative values of Clog P, suggest that the activity decreases with an increase hydrophobicity of the compounds. While the bi-parametric equations that are not involve Clog P descriptor gave very poor models with R^2 values for these models range between 0.00-0.29 as shown in Table 5.

Table 3: The values of descriptor and the observed and predicted activities (log 1/C) by Eq. 1

N0.	ClogP	Observed activity log 1/C	predicted activity log 1/C	Residual
1	4.79	6.59	6.653	-0.063
2	5.85	6.17	6.270	-0.100
3	5.48	6.32	6.404	-0.084
6	4.90	6.66	6.613	0.047
7	4.31	6.82	6.826	-0.006
8	3.94	7.05	6.960	0.090
9	5.40	6.60	6.433	0.167
10	4.88	6.48	6.620	-0.140
11	6.40	6.16	6.071	0.089

residual is the different between observed and predicted biological activity.

Table 4: The values of R^2 for bi-parametric equations

Descriptor	R2
C log p	0.87
M.w	0.29
Log S	0.003
HOMO	0.00
LUMO	0.00
ΔE	0.15
CMR	0.17

In the tri-parametric equation (Eq. 2-6) that included Clog P with other descriptors as (MW), log S, HOMO, LUMO, and calculated molar refractivity (CMR). Very good models with R2 values 0.879, 0.874, 0.91, 0.875 and 899 respectively. The predicted values for equations 2-6 are reported under Table 6-10.

log1/C = -0.38 (±0.19)Clog P +7.92 (±4.06)MW+8.17 (± 1.28).....(2)

n = 9, R^2= 0.879 , SE = 0.118 , F = 21.80

log1/C = -0.36 (±0.22)Clog P +1.28 (±5.26)log S + 8.39 (± 0.78).....(3)

n = 9, R^2= 0.874 , SE = 0.120 , F = 21.91

log1/C = -0.36 (±0.11)Clog P +0.34 (±0.53)HOMO + 11.26 (± 4.57).....(4)

n = 9, R^2= 0.910 , SE = 0.10 , F = 30.52

log1/C = -0.35 (±0.13)Clog P -0.10 (±1.4)LUMO + 8.24 (± 1.95).....(5)

n = 9, R^2= 0.875 , SE = 0.120 , F = 21.03

log1/C = -0.41 (±0.16)Clog P +6.86 (±0.13)CMR + 7.87 (± 1.17).....(6)

n = 9, R^2= 0.899 , SE = 0.10 , F = 26.99

The tri-parametric that involve ClogP with MW, log S, εHOMO, and CMR as descriptors (Eq.2-4 and 6) gave good models with R^2 value of 0.897, 0.874, 0.91 and 0.899 respectively. In these models the negative values of Clog P, and positive values of MW, log S, εHOMO and CMR suggest that the activity decreases with an increase hydrophobicity and decrease MW, log S, HOMO and CMR of the compounds respectively.

while the tri-parametric equations that is involve Clog P and εLUMO descriptors (Eq.5) also gave good models with R^2 value of 0.899. In this

model the negative values of Clog P and εLUMO, suggest that the activity decreases with an increase hydrophobicity and εLUMO of the compounds.

Table 5: The values of descriptors and the observed and predicted activities (log 1/C) by Eq. 2

N0.	MW	ClogP	Aobs	Apredd	Residual
1	397.26	4.79	6.59	6.627	-0.037
2	476.15	5.85	6.17	6.278	-0.010
3	411.28	5.48	6.32	6.370	-0.050
6	427.28	4.90	6.66	6.608	0.051
7	413.26	4.31	6.82	6.826	-0.006
8	427.28	3.94	7.05	6.980	0.069
9	441.31	5.40	6.60	6.425	0.1745
10	457.31	4.88	6.48	6.639	-0.159
11	511.40	6.40	6.16	6.092	0.0670

Table 6: The values of descriptors and the observed and predicted activities (log 1/C) by Eq.3

N0.	Log S	ClogP	A_{obs}[c]	Apredd	Residual
1	-6.606	4.79	6.59	6.654	-0.064
2	-7.424	5.85	6.17	6.269	-0.099
3	-6.957	5.48	6.32	6.404	-0.084
6	-6.656	4.90	6.66	6.614	0.045
7	-6449	4.31	6.82	6.819	1.99E
8	-6.134	3.94	7.05	6.962	0.0875
9	-6.863	5.40	6.60	6.433	0.166
10	-6.729	4.88	6.48	6.621	-0.141
11	-7.856	6.40	6.16	6.070	0.0895

Table 7: The values of descriptors and the observed and predicted activities (log 1/C) by Eq. 4

N0.	HOMO	ClogP	A_{obs}[c]	A_{pred}[d]	Residual
1	-8.54	4.79	6.59	6.631	-0.041
2	-8.68	5.85	6.17	6.200	-0.030
3	-8.46	5.48	6.32	6.409	-0.089
6	-8.36	4.90	6.66	6.653	-0.007
7	-8.37	4.31	6.82	6.861	-0.041
8	-8.641816	3.94	7.05	6.903	-0.147
9	-8.150362	5.40	6.60	6.544	-0.056
10	-8.553625	4.88	6.48	6.594	-0.114
11	-8.539686	6.40	6.16	6.051	-0.109

Table 8: The values of descriptors and the observed and predicted activities (log 1/C) by Eq.5

N0.	LUMO	ClogP	A_{obs}[c]	A_{pred}[d]	Residual
1	-1.137938	4.79	6.59	6.647	-0.057
2	-1.263135	5.85	6.17	6.280	-0.110
3	-1.108049	5.48	6.32	6.396	-0.076
6	-1.115858	4.90	6.66	6.605	0.054
7	-1.328826	4.31	6.82	6.840	-0.020
8	-1.182868	3.94	7.05	6.957	0.092
9	-1.223015	5.40	6.60	6.437	0.162
10	-1.147969	4.88	6.48	6.616	-0.136
11	-1.137938	6.40	6.16	6.069	0.090

Table 9: The values of descriptors and the observed and predicted activities (log 1/C) by Eq. 6

N0.	CMR	ClogP	A_{obs}[c]	A_{pred}[d]	Residual
1	10.7941	4.79	6.59	6.609	-0.019
2	11.5711	5.85	6.17	6.217	-0.047
3	11.2579	5.48	6.32	6.351	-0.0315
6	11.411	4.90	6.66	6.605	0.054
7	10.9472	4.31	6.82	6.821	-0.0012
8	11.411	3.94	7.05	7.008	0.0416
9	11.8748	5.40	6.60	6.427	0.1725
10	12.0279	4.88	6.48	6.656	-0.1761
11	13.9847	6.40	6.16	6.152	0.0073

Equation 2-6 indicates a strong dependency of the activity on the hydrophobicity of compounds, since the tri-parametric that not involve Clog P descriptor gave poor model with R^2 values between 0.00-0.25 as shown in Table 11. In addition, equation 4 give the best tri-parametric model with R^2 0.91, which greater than the R^2 of the other models in this study, therefore can be used as a model to design new inhibitor of Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase.

Table 10: The values of R^2 for tri-parametric equations

Descriptor 1	Descriptor 2	No. of compound	R2
Clog p	M.W	9	0.879
Clogp	Log S	9	0.874
Clog p	HOMO	9	0.91
Clog p	LUMO	9	0.875
Clog p	ΔE	9	0.905
Clog p	CMR	9	0.899
MW	Log S	9	0.23
MW	HOMO	9	0.18
MW	LUMO	9	0.25
MW	ΔE	9	0.20
MW	CMR	9	0.18
Log s	HOMO	9	0.00
Log S	LUMO	9	0.00
Log s	ΔE	9	0.00
Log s	CMR	9	0.09
HOMO	LUMO	9	0.00
HOMO	ΔE	9	0.00
HOMO	CMR	9	0.07
LUMO	ΔE	9	0.00
LUMO	CMR	9	0.05
ΔE	CMR	9	0.07

So our results indicated that both Clog P and εHOMO are of comparable significance on the regression equation goodness and when these two parameters were not involve in the model gave poor models.

CONCLUSION

The study indicated that QSAR of biological activity represented by log 1/C of pyrido[3,4-d] pyrimidine derivatives can be modeled with semi-emperical based quantum mechanical molecular descriptors. The comparison of all models indicated that the Clog P and εHOMO descriptors are more reliable than other descriptors in prediction of biological activity of pyrido[3,4-d] pyrimidine derivatives as Inhibitors of Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase.

References

1. An introduction to medicinal chemistry by Graham L Patric 3rd edition pagee no:271-298

2. Li, L., Qian, S., Alexander,K, N., Kennelh, F.., Chin-Chung, W., Ching-Yuan, S., and Kuo-Hsiung, L; ***Antitumor agents 250. Design and synthesis of new curcumin as potential antiprostat cancer agents***. J. Med. Chem; 49, 3963-3972, 2006.

3. Lorentz, J., and Sorana, B***; Molecular Descriptors Family on Structure Activity Relationships***.Leonardo Electronic Journal of Practices and Technologies;55-102, 2006.

4. Hari, S. K. P; ***Molecular Modeling Studies of Curcumin analogs as anti - angiogenic agents***; Thesis; 2008.

5. Cristina, D. M., Adina,C., Gabriel, K., and Mircea, V.D; ***Application to QSAR studies of 2-furylethylene derivatives***; J.Meth.Chem;1-10, 2008.

6. Shinde, M., Jangam, T., Chougule, U., Mahanthesh, M., and Bhatia, M; ***QSAR-a novel tool in drug design***; International Journal of Pharmaceutical Applications; 62-75, 2010.

7. Robert Jr R. Properties of FDA-approved small molecule protein kinase inhibitors: A 2020 update. Pharmacol. Res. 2020;152:104609.

8. Klutchko, S. R.; Hamby, J. M.; Boschelli, D. H.; Wu, Z.; Kraker, A. J.; Amar, A. M.; Hartl, B. G.; Shen, C.; Klohs, W. D.; Steinkampf, R. W.; Driscoll, D. L.; Nelson, J. M.; Elliott, W. L.; Roberts, B. J.; Stoner, C. L.; Vincent, P. W.; Dykes, D. J.; Panek, R. L.; Lu, G. H.; Major, T. C.; Dahring, T. K.; Hallak, H.; Bradford, L. A.; Showalter, H. D. H.; Doherty, A. M. 2-Substituted Aminopyrido[2,3-d]- pyrimidin-7(8H)-ones. Structure-activity Relationships Against Selected Tyrosine Kinases and in Vitro and in Vivo Anticancer Activity. J. Med. Chem. 1998, 41, 3276-3292.