Journal of Enzyme Inhibition and Medicinal Chemistry, 2009; 24(4): 937–948

informa healthcare

RESEARCH ARTICLE

QSAR of aminopyrido[2,3-d]pyrimidin-7-yl derivatives: Anticancer drug design by computed descriptors

Sisir Nandi and Manish C. Bagchi

Structural Biology and Bioinformatics Division, Indian Institute of Chemical Biology, 4 Raja S.C. Mullick Road, Jadavpur, Calcutta 700032, India

Abstract

A series of aminopyrido[2,3-d]pyrimidin-7-yl derivatives acting as potential tyrosine kinase inhibitors having anticancer activities have been considered in the present investigation for the quantitative structure-activity relationship studies based on 2D and 3D QSAR approaches. For this purpose, various theoretical molecular descriptors were computed solely from the structures of these compounds. As the number of molecular descriptors greatly exceeds the number of observations, conventional regression does not produce reliable models and therefore, ridge regression methodology was used to solve this problem. The influence of different classes of molecular descriptors on the activity has been predicted and the most significant descriptors were obtained using the ridge regression models. Partial least squares (PLS) models were developed based on the training set for the 3D QSAR models of the above compounds. The influences of steric and electrostatic field effects generated by the contribution plots are discussed.

Keywords: *QSAR; ridge regression; structural descriptors; murine tumors; pyrido pyrimidine derivatives; anticancer drug design; tyrosine kinase; inhibition*

Introduction

A considerable number of experimental investigations have been made over the past several years in searching for tyrosine kinase inhibitors which act as therapeutic potential anticancer agents. Hamby et al. [1] synthesized a series of pyrido[2,3-d]pyrimidine derivatives, acting as inhibitors of various tyrosine protein kinases broadly classified as receptor or non-receptor (cellular-Src, c-Src) types [2] and explored structure-activity relationships for these compounds. Klutchko and co-workers [3] synthesized 2-aminopyrido[2,3-*d*]pyrimidin-7(8*H*)-ones as a novel class of potent, broadly active tyrosine kinase (TK) inhibitors and tested their in vivo and in vitro anticancer activities. The synthesis of a number of pyrido[1′,2′:1,5]pyrazolo[3,4-*d*] pyrimidine derivatives by Showalter et al. [4] showed some interesting findings in anticancer effects. Schroeder et al. [5] synthesized numerous aminopyrido[2,3-d]pyrimidin-7-yl derivatives as the potential tyrosine kinase inhibitors and the anticancer activities of these compounds have been reported. The biochemical mechanisms of these compounds is to inhibit selected tyrosine kinases such as PDGFr, FGFr and c-Src at the ATP binding site and have been shown to exert growth delay effects on rapidly proliferating cell lines relating to murine tumors such as colon cancer, breast cancer, glioma and ovarion tumors [2]. In their attempt to study structure-activity relationships of aminopyrido[2,3-d]pyrimidin-7-yl derivatives, Schroeder et al. determined experimental anticancer activities of such compounds by introducing various substituents at different positions of the fused ring system. But there is hardly any quantitative structure-activity relationships study using non empirical parameters on the aminopyrido [2,3-d] pyrimidine-7-yl derivatives. QSAR studies where physical properties and physicochemical substituent constants are only used for the prediction of other more complex physicochemical, biomedicinal and toxicological properties cannot explain the model completely as the physicochemical properties of the compounds concerned are not always available. Hence, there is a need to develop QSAR models based on theoretical molecular descriptors, which can be calculated directly from the chemical structure of the compounds [6–11]. In this paper, an attempt has been made to formulate QSAR models from the standpoint of 2D and 3D approaches using calculated molecular descriptors that

Addresss for Correspondence: Dr M. C. Bagchi, Tel.: + 91 33 2473 3491/3493/0493/6793; Fax: + 91 33 2473 5197, + 91-33-2472 3967; E-mail: mcbagchi@iicb.res.in

(Received 11 July 2008; revised 03 September 2008; accepted 09 September 2008)

ISSN 1475-6366 print/ISSN 1475-6374 online © 2009 Informa UK Ltd
DOI: 10.1080/14756360802519327

http://www.informapharmascience.com/enz

could provide predictions of activities of such compounds, real or hypothetical.

One approach of treating 2D QSAR problems is to develop QSAR models based on a large set of theoretical molecular descriptors using ridge regression methodology [12] that ultimately predicts the influence of different class of calculated molecular descriptors on biological activities of the molecules considered for investigation. The purpose of three-dimensional quantitative structure-activity relationship (3D-QSAR) study is to find the specific contributions of steric and electrostatic effects on the biological activities of the inhibitors [13]. It is evident from this study that the application of 3D QSAR method generates QSAR models with improved internal and external prediction accuracy. One is led to the fact that such treatment provides better quality QSAR models by predicting biological activities of the new potential compounds of interest. It is believed that studies in this direction would reveal a useful approximation for the structural requirements in the design of more promising anticancer compounds, thus providing a better tool for rational drug design [6–11].

Materials and methods

Biological activity data of aminopyrido[2,3-d] pyrimidin-7-yl derivatives

The anticancer activities of aminopyrido[2,3-d]pyrimidin-7-yl derivatives for PDGFr, FGFr and c-Src kinase inhibition, have been reported by Schroeder et al. [5] by considering structural modifications with the introduction of different aliphatic and aromatic substituents at 2, 6 and 7 positions of the aminopyrido[2,3-d]pyrimidin-7-yl fused ring system. The biological activity data, in the form of IC_{50}, were determined experimentally [5]. All the biological activity values were converted in terms of pIC_{50} or $\log(1/IC_{50})$ where IC_{50} represents the concentration of these compounds that produce 50% kinase inhibition. These activities may be considered for the development of a valid QSAR model. QSAR models developed by using experimental properties as independent variables, are essentially property-property correlation, whereas models developed by using calculated molecular descriptors solely from the molecular structures of these aminopyrido[2,3-d]pyrimidin-7-yl compounds will give some insight on structure-property correlations. Our aim is to utilize these activity data for creating structure property correlations, which may provide a better tool for the rational drug design [14]. Table 1 shows the chemical structure of all compounds along with their biological activity values.

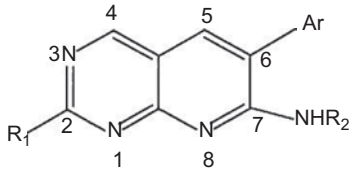
Structural modifications at 2, 6 and 7 positions of the aminopyrido[2,3-d]pyrimidin-7-yl fused ring system are important for the development of potent compounds with desired pharmacological activity. It is seen from Table 1 that C-6 aryl substituent specially 2, 6-dicholorophenyl substitution produces good potency of the compounds whereas introduction of tertiary butyl urea group at N-7 amino position of these compounds increases activity. Alkyl amino side chain at the C-2 position is essential for producing optimal kinase inhibition. Extension of carbon backbone of alkyl amino side chain at C-2 along with terminator as N-(4-methylpiperazino) or N,N-diethyl amino group, produces satisfactory kinase inhibitory activity.

Molecular descriptor calculation

Theoretical molecular descriptors are the numerical representation of molecules. It encodes chemical information of molecular structures. These descriptors are more useful in the area of theoretical molecular design and discovery research. A recent trend in this direction is the use of theoretical molecular descriptors, which can be calculated directly from molecular structure of the given compound using different level of molecular information ranging from 2D to 3D geometry [15]. For the development of 2D QSAR models, descriptors such as physicochemical, constitutional, electrostatic, geometrical, and topological indices have been used in the present investigation. The physicochemical descriptors include AlogP98 value, AMR value, buffer solubility, polarizability, vapour density, water solubility, solvation free energy, and so forth. Constitutional descriptors deals with composition of molecule, such as molecular weight, molecular formula, number of atoms, bonds, and rings, etc whereas electrostatic descriptors are based on the electronic and electrostatic structure of a molecule, for example, partial atomic charges, electronegetivity of the atoms, molecular electrostatic potential, etc. Geometrical descriptors are mainly the shape descriptors which are still more complex, encoding the three dimensional properties of molecules. Topological descriptors are the largest set of molecular descriptors, which may again be subdivided into two classes- topostructural, and topochemical descriptors. Topostructural descriptors encode information strictly on the neighbourhood and connectivity of atoms within the molecule, while the topochemical descriptors encode information relating to both the topology of the molecule and the chemical nature of atoms and bonds within it [8–11].

In our present study, the theoretical molecular descriptors have been calculated using PreADMET software package [16], which is a web-based application for predicting ADME data and building drug-like and lead-like compounds for high throughput screening (HTS) and library of combinatorial chemistry using in-silico method. This program can calculate about 955 molecular descriptors including physicochemical, constitutional, electrostatic, geometrical and topological indices. The input file may be created either by drawing the chemical structure or using an appropriate SMILES notation of the compound concerned. A total number of 495 molecular descriptors, useful for our purpose, were calculated using PreADMET program and prior to model development, descriptors with zero values or constant value for, or nearly all, of the compounds and that are completely correlated ($r = 1.0$) with another descriptor, are eliminated from the descriptor set. Table 2 represents the symbols of calculated molecular descriptors considered in our study.

Table 1. Structures of 2, 6, 7-substituted aminopyrido[2,3-d]pyrimidine-7-yl derivatives with activities.



Comp No.	Substituents			IC_{50} (μM)			pIC_{50} (μM)		
	Ar	R_1	R_2	PDGFr	FGFr	c-Src	PDGFr	FGFr	c-Src
1	2,6-$(cl)_2$Ph	NH_2	t-BuNHCO	1.2	0.14	0.22	−0.079*	0.853*	0.657*
2	2,6-$(cl)_2$Ph	NH_2	H	16.0	3.0	0.21	−1.204	−0.477	0.677**
3	2,6-$(cl)_2$Ph	$NH(CH_2)_3NEt_2$	H	46.0	2.4	0.75	−1.662	−0.380	0.124**
4	2,6-$(cl)_2$Ph	$NH(CH_2)_3NEt_2$	t-BuNHCO	0.66	0.082	0.073	0.180*	1.086**	1.136**
5	2,6-$(Me)_2$Ph	NH_2	H	29.0	13.0	0.43	−1.462	−1.113	0.366*
6	2,6-$(Me)_2$Ph	NH_2	t-BuNHCO	0.34	0.40	0.11	0.468*	0.397*	0.958*
7	2,6-$(Me)_2$Ph	$NH(CH_2)_3NEt_2$	H	25.0	18.0	0.45	−1.397	−1.255	0.346**
8	2,6-$(Me)_2$Ph	$NH(CH_2)_3NEt_2$	t-BuNHCO	0.80	0.34	0.098	0.096*	0.468*	1.008*
9	2,6-$(Br)_2$Ph	NH_2	H	-	13.0	1.6	-	−1.113	−0.204
10	2,6-$(Br)_2$Ph	NH_2	t-BuNHCO	1.4	0.29	0.21	−0.146**	0.537**	0.677**
11	2,6-$(Br)_2$Ph	$NH(CH_2)_3NEt_2$	H	-	8.3	0.76	-	−0.919	0.119*
12	2,6-$(Br)_2$Ph	$NH(CH_2)_3NEt_2$	t-BuNHCO	1.1	0.19	0.097	−0.041**	0.721**	1.013**
13	2,3,5,6-$(Me)_4$Ph	NH_2	t-BuNHCO	-	0.71	-	-	0.148**	-
14	3,5-$(MeO)_2$Ph	NH_2	H	-	0.23	-	-	0.638*	-
15	3,5-$(MeO)_2$Ph	NH_2	t-BuNHCO	-	0.048	-	-	1.318*	-
16	2,6-$(cl)_2$Ph	$NH(CH_2)_2NEt_2$	EtNHCO	12.0	1.3	2.6	−1.079	−0.113	−0.414
17	2,6-$(cl)_2$Ph	$NH(CH_2)_2NEt_2$	t-BuNHCO	9.0	1.8	4.1	−0.954	−0.255	−0.612
18	2,6-$(cl)_2$Ph	$NH(CH_2)_3NEt_2$	EtNHCO	1.3	0.13	0.094	−0.113**	0.886**	1.026**
19	2,6-$(cl)_2$Ph	$NH(CH_2)_3NEt_2$	i-PrNHCO	1.1	0.077	0.078	−0.041*	1.113*	1.107*
20	2,6-$(cl)_2$Ph	$NH(CH_2)_4NEt_2$	EtNHCO	0.21	0.049	0.018	0.677*	1.309*	1.744**
21	2,6-$(cl)_2$Ph	$NH(CH_2)_4NEt_2$	t-BuNHCO	0.36	0.048	0.0074	0.443*	1.318*	2.130*
22	2,6-$(cl)_2$Ph	$NH(CH_2)_4NEt_2$	cyclohexylNHCO	0.33	0.043	0.012	0.481*	1.366**	1.920**
23	2,6-$(cl)_2$Ph	$NH(CH_2)_4NEt_2$	PhNHCO	0.45	0.11	0.0075	0.346**	0.958**	2.124*
24	2,6-$(cl)_2$Ph	$NH(CH_2)_3NMe_2$	t-BuNHCO	0.68	0.075	0.12	0.167*	1.124*	0.920**
25	2,6-$(cl)_2$Ph	$NMe(CH_2)_3NMe_2$	t-BuNHCO	16.0	2.0	1.1	−1.204	−0.301	−0.041
26	2,6-$(cl)_2$Ph	$NHCH_2CMe_2CH_2NMe_2$	t-BuNHCO	3.2	0.21	3.5	−0.505	0.677	−0.544
27	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(morpholin-1-yl)	t-BuNHCO	0.84	0.072	0.10	0.075*	1.142*	1.000*
28	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(2-methylpiperidin-1-yl)	t-BuNHCO	0.73	0.06	0.016	0.136**	1.221**	1.795*
29	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	H	9.6	0.45	0.18	−0.982	0.346*	0.744*
30	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	t-BuNHCO	0.47	0.051	0.032	0.327**	1.292*	1.494**
31	2,6-$(cl)_2$Ph	$NH(CH_2)_4$(N-methylpiperazin-1-yl)	t-BuNHCO	0.28	0.035	0.010	0.552*	1.455*	−2.000*
32	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	H_2NCO	2.4	0.14	0.061	−0.380	0.853*	1.214**
33	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	EtNHCO	0.42	0.053	0.024	0.376*	1.275*	1.619*
34	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	allylNHCO	0.76	0.035	0.022	0.119*	1.455**	1.657**
35	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	i-PrNHCO	0.55	0.034	0.019	0.259*	1.468**	1.721**
36	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	n-octylNHCO	6.2	0.42	0.12	−0.792	0.376*	0.920*
37	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	benzylNHCO	2.5	0.062	0.030	−0.397	1.207*	1.522**
38	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	cyclohexylNHCO	0.37	0.029	0.031	0.431	1.537*	1.508*
39	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	adamantylNHCO	2.8	0.12	0.12	−0.447	0.920**	0.920**
40	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	$BOCNH(CH_2)_2NHCO$	4.2	0.067	0.15	−0.623	1.173*	0.823*
41	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	$Me_2N(CH_2)_2NHCO$	2.4	0.075	0.02	−0.380	1.124*	1.698**
42	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	Et_2NCO	-	5.5	-	-	−0.740	-
43	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	PhNHCO	0.57	0.084	0.015	0.244*	1.075*	1.823*
44	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	4-ClPhNHCO	1.8	0.12	0.04	−0.255	0.920*	0.397*
45	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	4-BrPhNHCO	1.5	0.11	0.033	−0.176	0.958*	1.481*
46	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	4-CF_3PhNHCO	5.5	0.61	0.19	0.740	0.214*	0.721*
47	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	3,4-$(Cl)_2$PhNHCO	5.0	0.60	0.092	−0.698	0.221*	1.026*
48	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	4-MePhNHCO	0.84	0.11	0.036	0.075*	0.958*	1.443*
49	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	2-MeOPhNHCO	0.66	0.067	0.016	0.180	1.173*	1.795*
50	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	3-MeOPhNHCO	0.91	0.063	0.029	0.040*	1.200**	1.537*

Table 1. Continued on next page

Table 1. Continued.

Comp No.	Substituents			IC_{50} (μM)			pIC_{50} (μM)		
	Ar	R_1	R_2	PDGFr	FGFr	c-Src	PDGFr	FGFr	c-Src
51	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	4-MeOPhNHCO	0.68	0.074	0.033	0.167*	1.130*	1.481*
52	2,6-$(cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	1-naphthylNHCO	2.6	0.30	0.13	−0.414	0.522*	0.886*
53	2,6-$(Cl)_2$Ph	$NH(CH_2)_3NEt_2$	t-$BuCH_2CO$	-	27.0	19.0	-	−1.143	−1.278
54	2,6-$(Cl)_2$Ph	$NH(CH_2)_3NEt_2$	Me_2NCH	-	10.0	4.0	-	−1.000	−0.602
55	2,6-$(Cl)_2$Ph	$NH(CH_2)_3NEt_2$	EtNHCS	5.0	0.26	0.13	−0.698	0.585*	0.886*
56	2,6-$(Cl)_2$Ph	$NH(CH_2)_3NEt_2$	EtNHCNH	-	2.7	1.4	-	−0.431	−0.146
57	2,6-$(Cl)_2$Ph	$NH(CH_2)_4NEt_2$	(morpholyn-1-yl) $(CH_2)_3$NHCS	1.1	0.13	0.022	−0.041	0.886*	1.657*
58	2,6-$(Cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	EtCH_2CO	-	3.7	2.7	-	−0.568	−0.431
59	2,6-$(Cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	t-$BuCH_2CO$	-	8.0	8.4	-	−0.903	−0.924
60	2,6-$(Cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	PhCH_2CO	-	6.8	3.3	-	−0.832	−0.518
61	2,6-$(Cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	$PhSO_2$	6.9	0.068	0.022	0.838	1.167*	1.657*
62	2,6-$(Cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	PhNHCNH	37.0	2.8	0.86	−1.568	−0.447	0.065
63	2,6-$(Cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	i-PrNHCN-i-Pr	-	2.5	0.85	-	−0.397	0.070
64	2,6-$(Cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	t-BuNHCS	1.9	0.10	0.22	−0.278	1.000*	0.657*
65	2,6-$(Cl)_2$Ph	$NH(CH_2)_3$(N-methylpiperazin-1-yl)	PhNHCS	2.2	0.32	0.022	−0.342	0.494*	1.657*

* Compounds considered in 3D QSAR analysis

** Compounds belonging to test set in 3D QSAR analysis

For the purpose of 3D QSAR analysis, chemical structures of all the compounds have been drawn using the builder module in Molecular Design Suite (MDS) 3.5 software package [17]. Molecular descriptors such as steric and electrostatic fields were calculated utilizing MDS 3.5 software which produces more than 6000 descriptors and prior to model development descriptors having zero values or same values were removed.

Ridge regression: A model building method in 2D QSAR

QSAR models have been generated utilizing different types of theoretical molecular descriptors such as physicochemical, constitutional and geometrical, electrostatic and topological which can be calculated solely from the structures of various sets of aminopyrido[2,3-d]pyrimidin-7-yl derivatives. Consideration of theoretical molecular descriptors has often proved to be more powerful tools having wide applications in quantitative structure-activity relationships modeling [18]. To establish correlations between biological activity and structural or property descriptors of the different sets of aminopyrido[2,3-d]pyrimidin-7-yl derivatives, it is essential to develop a regression or an input-output model. Multivariate regression analysis (MRA), one of the oldest data reduction methodologies, continues to be widely used in QSAR [19] as it does not impose any restriction on the type and number of graphical invariants used in structure-property- activity studies. For a valid statistical significance of the MRA, it is necessary to restrict the maximal number of descriptors, which will depend on the number of compounds investigated [20,21]. In order to avoid ambiguities in the interpretation of regression, only few parameters, or ideally a single parameter may be used. Conventional regression (OLS) does not produce reliable models when the number of descriptors exceeds the number of observations [22,23]. The commonly used alternative regression methodologies include partial least squares (PLS) [24,25], principal components regression (PCR) [26] and ridge regression (RR) [12]. All these methods except OLS are intended to work when the independent variables are highly multicollinear and when the number of independent variables is substantially greater than the number of observations. In addition, these regression methods make use of entire available pool of independent variables as opposed to subset regression which introduces bias and may result in the elimination of important parameters, leading to dramatic reductions in the variance of the estimated model coefficients. Formal comparisons have consistently shown subsetting to be less effective than the alternative methods, such as these, that retain all of the independent variables and use other approaches to deal with the rank deficiency [27]. In ridge regression, descriptors are transformed into principal components (PCs) and these PCs are used as new predictors. Unlike PCR, RR retains all principal components and shrinks them differentially according to their eigen values. In PCR, descriptors are transformed into principal components after which a subset of PCs is used in ordinary least square regression, whereas, PLS are also related to principal component network, but the shrinkage is not based on Eigen values. Statistical theory suggests that RR is the best of the three methods and this has been considered by us for developing 2D QSAR models. The RR vector of regression coefficients, **b**, is given by:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X} + k\mathbf{I})^{-1}\,\mathbf{X}^T\mathbf{Y}$$

Where **X** is the matrix of descriptors, **Y** is the vector of observed activities, **I** is an identity matrix, and k is a non-negative constant known as the "ridge" constant. If k = 0, RR reduces to conventional OLS regression.

Further model building methods in 3D QSAR

Geometry Optimization

Three dimensional quantitative structure-activity relationship studies of aminopyrido[2,3-d]pyrimidin-7-yl

Table 2. List of calculated molecular descriptors used in 2D QSAR analysis.

Descriptor Classes	Descriptor Names
Constitutional Descriptors	No. amino groups primary, No. amino groups tertiary, No. amino groups secondary, No. amide groups, No. ester groups, No. halogen atoms, Molecular weight, Formal charge, No. Total atoms, No. Rotatable bonds, Fraction of Rotatable bonds, No. Rigid bonds, No. Rings, No. single bonds, No. double bonds, No. H-bond acceptors, No. H-bond donors, Ratio donors to acceptor, No. Aromatic rings, No. aromatic bonds, No. positive charged groups, No. positive chargable groups.
Geometrical Descriptors	2D-VDW surface, 2D-VDW volume, 2D-VSA hydrophobic, Fraction of 2D-VSA hydrophobic, 2D-VSA hydrophobic_sat, 2D-VSA hydrophobic_unsat, 2D-VSA other, 2D-VSA polar, Fraction of 2D-VSA polar, 2D-VSA Hbond acceptor, 2D-VSA Hbond donor, 2D-VSA Hbond all, Fraction of 2D-VSA Hbond, Topological PSA, 2D-VSA positive chargable groups, Fraction of 2D-VSA chargable groups, 2D-VSA positive charged groups, 2D-VSA chargable groups.
Electrostatic Descriptors	Max positive charge, Max positive hydrogen charge, Total negative charge, Total absolute atomic charge, Local dipole index, Relative negative charge, PNSA1 (Partial Positive Surface Area 1st type), PPSA2, PNSA3, DPSA2 (Difference in Charged Partial Surface Area), FPSA1 (Fractional charged partial positive surface area 1st type), FPSA3, FNSA2, WPSA1 (Surface weighted charged partial positive surface area 1st type),WPSA3, WNSA2 (Surface weighted charged partial negative surface area 2nd type), RPCS (Relative positive charge surface area), Hydrophobic SA – MPEOE, Negative charged polar SA – MPEOE, SADH2 (Surface area on donor hydrogens 2nd type), CHDH1 (Charge on donatable hydrogens 1st type), CHDH3, SCDH2 (Surface weighted charged area on donor hydrogens 2nd type), SAAA1 (Surface weighted charged area on acceptor atoms 1st type), SAAA3, CHAA2 (Charge on acceptors atoms 2nd type), SCAA1 SCAA1 (Surface weighted charged area on acceptor atoms 1st type), SCAA3, HRPCG, HRNCS, HRNCG, Max negative charge, Total positive charge, Charge polarization, Polarity parameter, Relative positive charge, PPSA1, PPSA3, PNSA2, DPSA1, DPSA3, FPSA2, FNSA1, FNSA3, WPSA2, WNSA1, WNSA3, RNCS, Positive charged polar SA – MPEOE, SADH1 (Surface area on donor hydrogens 1st type), SADH3, CHDH2, SCDH1, SCDH3, SAAA2, CHAA1, CHAA3, SCAA2, ACGD, HRPCS, CHGD.
Topological Descriptors	Total structure connectivity index, Chi 0 (Simple zero order chi index), Chi 1, Chi 2, Chi 3 path (Simple third order path chi index), Chi 3 cluster (Simple 3rd order cluster chi index), Chi 4 cluster, Chi 4 path, Chi 5 path, Chi 4 path/cluster (Simple 4th order path/cluster chi index), VChi 0 (Valance zero order chi index), VChi 1, VChi 2, VChi 3 path (Valance 3rd order path chi index), VChi 4 path, VChi 3 cluster, VChi 4 cluster, VChi 4 path/cluster, VChi 5 path, Kier shape 1 (encodes the degree of cyclicity in the graph, decreases as graph cyclicity increases), Kier shape 3 (encodes the degree of separated branching in the graph,increases as the degree of separation in branching increases.), Kier alpha 1 (1st Order Kappa Alpha Shape Index), Kier alpha 2, Kier alpha 3, Kier flexibility, Kier symmetry index, Delta Chi 0 (Delta zero order chi index), Delta Chi 1, Delta Chi 2, Delta Chi 3 path, Delta Chi 3 cluster, Delta Chi 4 path, Delta Chi 4 cluster, Delta Chi 5 path, Delta Chi 4 path/cluster, Difference chi 0 (Difference simple zero order chi index), Difference chi 1, Difference chi 2, Difference chi 3, Difference chi 4, Difference chi 5, IC (information content index), BIC (bond information content), CIC (complementary information content), SIC (structural information content), IAC total (total information index of atomic composition), I_adj_equ (Information index based on the vertex adjacency matrix equality), I_adj_mag (Information index based on the vertex adjacency matrix magnitude), I_adj_deg_equ (Information index based on the degree adjacency matrix equality), I_adj_deg_mag, I_dist_equ (Information index based on the distance matrix equality), I_dist_mag (Information index based on the distance matrix magnitude), I_edge_adj_equ (Information index based on the edge adjacency matrix equality), I_edge_adj_mag (Information index based on the edge adjacency matrix magnitude), I_edge_adj_deg_equ, I_edge_adj_deg_mag, I_edge_dist_equ, I_edge_dist_mag, Wiener index (Half-sum of the off-diagonal elements of the distance matrix of a graph), Hyper Wiener index, Harary index (Half-sum of the off-diagonal elements of the reciprocal molecular distance matrix), 1st Zagreb (1st Zegreb index), 2nd Zagreb, Quadratic index, Rouvray index, 2-MTI (Schultz Molecular Topological Index (MTI)), 2-MTI prime (Schultz MTI by valence vertex degrees), Gutman MTI, Graph diameter, Graph radius, Graph Petitjean, Eccentric connectivity index, Eccentric adjacency index, Platt number, Odd-even index, Vertex degree-distance index, Ring degree-distance index, Balaban index JX, Balaban index JY, Xu (Xu index), Superpendentic index, Unipolarity_distance_matrix, Centralization_distance_matrix, Dispersion_distance_matrix, SC-0 (Subgraph Count Index of order 0), SC-1, SC-2, SC-3 path, SC-3 cluster, SC-4 path, Solvation chi 4 path/cluster, Solvation chi 5 path, VS-0 (Valence Shell Count of order 0), VS-1, VS-2, VS-3, VS-4, VS-5, Molecular walk count 2, Molecular walk count 3, Molecular walk count 4, Molecular walk count 5, Path/walk 2, Path/walk 3, Path/walk 4, Path/walk 5, Narumi ATI (Narumi simple topological index (log)), Narumi HTI (Narumi harmonic topological index), Narumi GTI (Narumi geometric topological index), Pogliani index, Ramification index, Degree complexity, Graph vertex complexity, Graph distance complexity, Graph distance index, Mean square distance index, Mean distance deviation, Edge Wiener index, Edge Hyper Wiener index, Edge MTI, Edge Gutman MTI, Edge connectivity index, E-state SsCH3, E-state SssCH2, E-state SdsCH, E-state SsssCH, E-state SdO, E-state S_hydrophobic, E-state S_hydrophobic_unsat, E-state S_polar, E-state S_hbond_donor, E-state SHdsCH, E-state SHCHnX, E-state SH_hydrophobic, E-state SdssC, E-state SsssN, E-state S_hydrophobic_sat, E-state S_none, E-state S_hbond_acceptor. E-state SsNH2, E-state SssNH, E-state SssO, E-state SsCl, E-state SsBr, E-state SHsNH2, E-state SHssNH, E-state SHCsats, E-state SHCsatu, , E-state SH_hbond_donor, E-state SaasC, E-state SssssC, E-state SdNH, E-state SaaN, E-state SdSr, E-state SHdNH, E-state SH_polar, E-state SH_positive_charged_group, E-state SaaCH, E-state SaaaC, E-state S_positive_charged_group, E-state SdsN, E-state SHaaCH, SC-4 cluster, SC-5 path, SC-7 path, SC-9 path, Solvation chi 0, Solvation chi 2, Solvation chi 3, Solvation chi 4 cluster, Cluster, SC-4 path/cluster, SC-6 path, SC-8 path, SC-10 path, Solvation chi 1, Solvation chi 3 path, Solvation chi 4 path.
Physicochemical Descriptors	Polarizability_Miller, SKlogP value, SKlogS value, SKlogPvp, SKlogS_buffer,SK_BP, AlogP98 value, Solvation Free Energy, SKlogD value, Water solubility, Vapor pressure, Buffer solubility, SK_MP, AMR value, AlogP98 002 C, AlogP98 004 C, AlogP98 006 C, AlogP98 008 C, AlogP98 024 C, AlogP98 026 C, AlogP98 040 C, AlogP98 047 H, AlogP98 049 H, AlogP98 051 H, AlogP98 053 H, AlogP98 059 O, AlogP98 067 N, AlogP98 069 N, AlogP98 094 Br, AlogP98 001 C, AlogP98 003 C, AlogP98 005 C, AlogP98 011 C, AlogP98 025 C, AlogP98 041 C, AlogP98 046 H, AlogP98 050 H, AlogP98 052 H, AlogP98 060 O, AlogP98 068 N, AlogP98 070 N, AlogP98 072 N, AlogP98 074 N, AlogP98 089 Cl, AlogP98 108 S.

derivatives were performed using Molecular Design Suite software version 3.5 [17]. For the purpose of our 3D QSARs, the compounds having higher inhibitory activity against three tyrosine kinases were selected and it has been indicated by single asterisk in Table 1. The three-dimensional structures of all selected compounds were constructed using the builder module in MDS 3.5 and their geometries were subsequently optimized using the Merck Molecular force field (MMFF) and MMFF charge [28]. Energy minimizations were performed considering a dielectric constant of 1.0 and the convergence criterion is 0.01 Kcal/mol.

Alignment of molecules

In 3D QSAR analysis, the most crucial step is the molecular alignment. This method is based on the conformational flexibility of molecules, thus representing the relative orientation of molecules in 3D space. The molecular alignment utility can be used to study the shape variation with respect to the base structure selected for alignment. This is an attempt to identify the best overlapping between the structures of molecules. In the present study, the molecules of the dataset were aligned by template based method [13] where a template structure is defined and used as a basis for alignment of a set of molecules and a reference molecule was chosen on which the other molecules of the align dataset gets align considering the chosen template. The reference molecule is chosen in such a way that the molecule should be in the most stable and bioactive conformation state among the series of molecules considered. The following Figure 1 is used as the template for alignment by considering common elements of the series.

The molecules such as 34, 15 and 65 were chosen for reference in PDGFr, FGFr and c-Src cases respectively. After optimizing, the template structure and reference molecule are used to superimpose all molecules from the series in three cases to obtain optimal alignment between the molecular structures necessary for ligand-receptor interactions [29]. The superimpositions of all molecules based on minimizing RMSD in the software in three cases are shown in Figures 2, 3 and 4.

The scope of the present investigation lies in considering all the above superimposed conformations of aminopyrido[2,3-d]pyrimidin-7-yl derivatives in the three cases. These conformations are presumed to be the biologically active structures, overlaid in their common binding mode. Each conformation is taken in turn, and the molecular fields around it are calculated. This is achieved by the generation of three-dimensional rectangular grid around the set of aligned molecules. Steric and electrostatic field interaction energies around the aligned molecules were

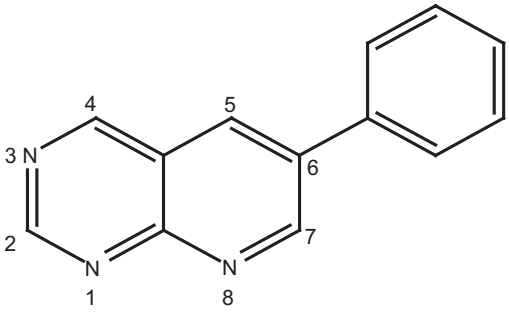


Figure 1. 6-Phenyl pyrido[2,3-d]pyrimidine.

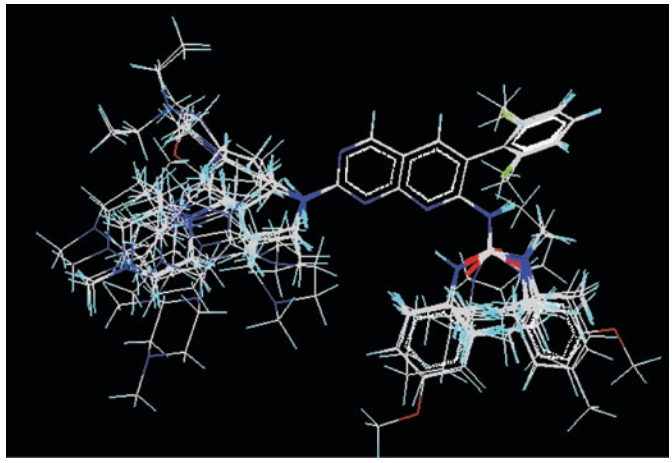


Figure 2. 3D view of aligned molecules for PDGFr.

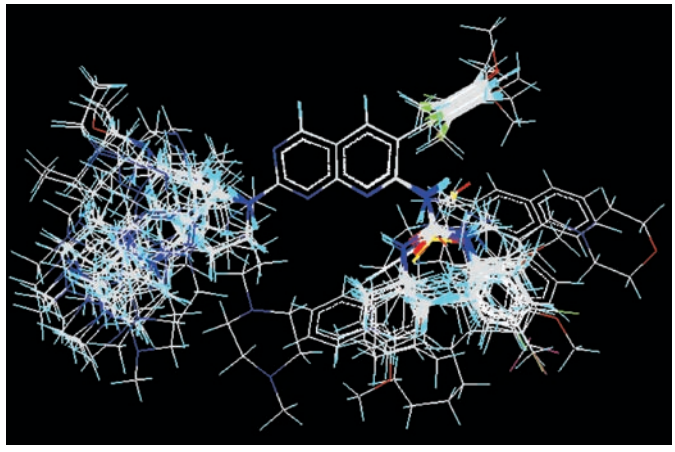


Figure 3. 3D view of aligned molecules for FGFr.

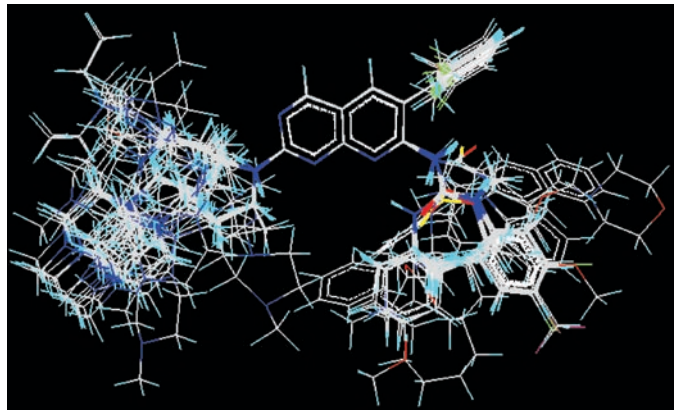


Figure 4. 3D view of aligned molecules for c-Src.

calculated at each grid point considering methyl probe of charge +1 with 10.0 kcal/mole electrostatic and 30.0 kcal/mole steric cut-off and the charge type of Gasteiger-Marsili [30] was set. A value of 1.0 was assigned to the dielectric constant. These electrostatic and steric fields are used as descriptors for 3D QSAR analysis.

Partial least squares (PLS)

Partial Least Squares methodology was implemented to derive 3D QSAR models of aminopyrido[2,3-d]pyrimidin-7-yl derivatives for PDGFr, FGFr and c-Src kinase inhibitors. Steric and electrostatic field descriptors were calculated and used as independent variables and pIC_{50} values of the compounds were considered as dependent variable in the PLS regression analysis. Internal validations of the models in all the three cases were made in terms of cross-validated q^2 and external predictability of the developed models were performed by calculating predictive r^2 (Pred_r^2) using following equations [31,32].

$$q^2 = 1 - \frac{\text{PRESS}}{\sum_{i=1}^{N}(y_i - \hat{y})^2} = 1 - \frac{\sum_{i=1}^{N}(y_i - y_{pred,i})^2}{\sum_{i=1}^{N}(y_i - \hat{y})^2} \quad (1)$$

$$\text{PRESS} = \sum_{i=1}^{N}(y_i - y_{pred,i})^2 \quad (2)$$

Where, PRESS denotes predicted sum of squared deviations between the observed activities (y_i) and predicted activities ($y_{pred,i}$) of the i-th molecules in the training set whereas ($\hat{y}$) is the mean of observed activities of all molecules in the training set.

$$\text{Pred\_}r^2 = 1 - \frac{\text{PRESS}}{\text{SSD}} = 1 - \frac{\sum_{i=1}^{N}(y_i - y_{pred,i})^2}{\sum_{i=1}^{N}(y_i - \hat{y})^2} \quad (3)$$

Where, y_i and $y_{pred,i}$ are the observed activities and predicted activities of i-th molecules in the test set.

$$\text{SSD} = \sum_{i=1}^{N}(y_i - \hat{y})^2 \quad (4)$$

SSD indicates sum of squared deviations between the observed activities (yi) of i-th molecules in the test set and mean activities ($\hat{y}$) of all molecules in the training set.

Results and discussion

2D QSAR results

An attempt has been made to generate 2D QSAR models based on computed molecular descriptors calculated solely from the structures of aminopyrido[2,3-d]pyrimidin-7-yl derivatives using ridge regression (RR) methodology. The NCSS software package [33] was used for the RR analysis. Ridge regression models demonstrate the influence pattern of various types of theoretical molecular descriptors viz. constitutional and geometrical, electrostatic, topological and physicochemical on the activities of aminopyrido[2,3-d]pyrimidin-7-yl derivatives. Table 3 provides the regression summary for QSARs of three sets of aminopyrido[2,3-d]pyrimidin-7-yl derivatives considered in the present investigation.

In case of PDGFr and FGFr, physicochemical descriptors contribute maximum influences having R^2 values of 0.956 and 0.946 respectively. This is followed by topological, electrostatic and constitutional and geometrical descriptors. From the table it is seen that topological descriptors provide a significant R^2 value of 0.912 and 0.940 for the above two kinase inhibition whereas electrostatic descriptors produce an R^2 values of 0.862 and 0.861. But constitutional and geometrical descriptors produce much inferior models having R^2 values of 0.687 and 0.750 for the first two cases. In considering RR models for c-Src, contribution of topological indices is maximum having a R^2 value of 0.768. The domination patterns of physicochemical and electrostatic descriptors over constitutional and geometrical descriptors are continued. Therefore, the ridge regression models generated by computed molecular descriptors can provide a good quality predictive model for the aminopyrido[2,3-d]pyrimidin-7-yl derivatives.

Another important statistical metric is the t value associated with each model, defined as the descriptor coefficient divided by its standard error [34]. Descriptors with large |t| values are important in the predictive model and, as such, can be examined in order to gain some understanding of the nature of property or activity of interest. Tables 4–6 indicate significant descriptors responsible for the good model predictions based on t value in case of PDGFr, FGFr and c-Src, respectively.

3D QSAR results

In the present paper, three dimensional quantitative structure-activity relationship studies of aminopyrido[2,3-d]pyrimidin-7-yl derivatives having inhibitory activities against PDGFr, FGFr and c-Src kinases, have been performed by applying stepwise variable selection method coupled with PLS model building technique. For the purpose, total data set was divided into training and test

Table 3. Regression summary for QSAR models of aminopyrido[2,3-d]pyrimidine-7-yl derivatives.

Molecular Descriptors	R^2 (Ridge regression)		
	PDGFr (N = 52)	FGFr (N = 65)	c-Src (N = 61)
(i) Constitutional and Geometrical	0.687	0.750	0.560
(ii) Electrostatic	0.862	0.861	0.713
(iii) Topological	0.912	0.940	0.768
(iv) Physicochemical Descriptors	0.956	0.946	0.759

Table 4. Significant descriptors responsible for the good model predictions based on t value in case of PDGFr.

Descriptor Classes	Significant Descriptors	t-value
Constitutional and Geometrical	Molecular weight	112.404
	2D van der Waals surface area (2D-VSA) hydrophobic sat	−57.988
	2D van der Waals(2D-VDW) surface	−54.183
	No. total atoms	48.063
	No. rigid bonds	−38.582
	2D-VDW volume	−35.660
Electrostatic	(WPSA2) Surface weighted charged partial positive surface area 2nd type	−676.189
	DPSA2(Difference in Charged Partial Surface Area)	668.014
	PPSA2(Partial Positive Surface Area)	247.150
	PNSA2(Partial Negative Surface Area)	−243.327
Physicochemical	SK_BP (added models of boiling point)	783.935
	SK_MP (added models of melting point)	−159.166
	AMR (calculated molecular refractivity index)	79.612
	Buffer solubility	45.434
	Solvation Free energy	−40.675
Topological	Edge Hyper Wiener Index	−67798.454
	Edge Gutman MTI	39032.406
	Graph Distance Index	−24144.376
	Weiner Index	1511.972

sets using sphere exclusion algorithm which allows constructing training set covering all descriptor space areas occupied by representative points and this method indicates diversity of sampling procedure [35]. Compounds with double asterisks in Table 1 were selected as test molecules for further model validation. The model quality for the training set was determined by calculating r^2, r^2se, cross-validated q^2 and q^2se whereas external validation in terms of predictive r^2 (pred_r^2) was done for the test set of compounds.

In case of PDGFr, PLS model for steric and electrostatic fields in the training set was obtained as follows:

$$\begin{aligned}\text{Activity (PDGFr)} = {} & 0.690 + (0.023)\,\text{E\_1276} \\ & + (0.017)\,\text{S\_1368} + (0.054)\,\text{E\_1255} \\ & + (0.010)\,\text{S\_797} - (0.011)\,\text{E\_1516}.\end{aligned} \quad (5)$$

n = 18, df = 14, r^2 = 0.936, q^2 = 0.866, F test = 67.909, r^2 se = 0.059, q^2 se = 0.086, no. of optimum components = 3

Similarly, 3D QSAR models using PLS regression for FGFr and c-Src were developed with the following statistical parameters.

$$\begin{aligned}\text{Activity (FGFr)} = {} & 0.592 + (0.120)\,\text{E\_802} - (0.063)\,\text{E\_1097} \\ & + (0.110)\,\text{E\_1074} + (0.080)\,\text{E\_1812} \\ & + (0.075)\,\text{E\_2243} + (0.031)\,\text{E\_972}.\end{aligned} \quad (6)$$

n = 34, df = 31, r^2 = 0.812, q^2 = 0.764, F test = 66.914, r^2 se = 0.166, q^2 se = 0.185, no. of optimum components = 2

Table 5. Significant descriptors responsible for the good model predictions based on t value in case of FGFr.

Descriptor Classes	Significant Descriptors	t-value
Constitutional and Geometrical	2D-VDW volume	250.327
	2D-VSA Hydrophobic	−211.353
	Molecular weight	−162.131
Electrostatic	DPSA2	598.581
	PPSA2	565.670
	WPSA1	−64.582
	RPCS(Relative positive charge surface area)	62.952
Physicochemical	SK_BP	−75.313
	SK_MP	−48.297
	Buffer solubility	40.836
	AMR value	32.921
Topological	Edge Hyper Wiener Index	−45608.666
	Hyper Wiener Index	−23163.730
	Gutman MTI	−21242.028
	Edge Gutman MTI	16498.756
	Graph Distance Index	−11977.933
	Edge MTI	10357.197

Table 6. Significant descriptors responsible for the good model predictions based on t value in case of c-Src.

Descriptor Classes	Significant Descriptors	t-value
Constitutional and Geometrical	2D-V2D-VSA Hydrophobic	−193.365
	SA Hydrophobic sat	−187.026
	2D-VDW surface	−184.764
	2D-VDW volume	130.961
Electrostatic	DPSA2	396.122
	PPSA2	361.811
	WPSA1	−337.542
Physicochemical	SK_MP	−82.067
	Buffer solubility	47.261
	SK_BP	−20.049
Topological	Edge Gutman MTI	−84542.709
	Edge Hyper Wiener Index	−60777.689
	Edge MTI	−41342.410
	Hyper Wiener Index	−30509.071

$$\begin{aligned}\text{Activity (c-Src)} = {} & 0.400 - (0.111)\,\text{S\_604} + (0.016)\,\text{S\_1798} \\ & + (0.031)\,\text{S\_1952} + (0.046)\,\text{E\_1974} \\ & + (0.025)\,\text{S\_1112} - (0.014)\,\text{S\_1808}\end{aligned} \quad (7)$$

n = 31, df = 28, r^2 = 0.853, q^2 = 0.804, F test = 81.527, r^2 se = 0.307, q^2 se = 0.355, no. of optimum components = 2

Here, n represents number of observations, df is the degree of freedom, r is the square root of the multiple R-square for regression, q^2 is the cross-validated r^2, F is the F-statistics for the regression model, and se is the corresponding standard error estimation.

The results indicate that 3D QSAR models for PDGFr, FGFr and c-Src generates 86.6%, 76.4% and 80.4% internal model prediction respectively.

The above QSAR models were validated on the compounds forming test sets and biological activity of test molecules were predicted. Tables 7–9 represent predicted activities of

Table 7. 3D QSAR derived predicted activities for PDGFr.

Test molecules	Observed Activity	Predicted Activity
10	−0.146	−0.106
12	−0.041	0.067
18	−0.113	−0.033
23	0.346	0.228
28	0.136	0.134
30	0.327	0.081

Table 8. 3D QSAR derived predicted activities for FGFr.

Test molecules	Observed Activity	Predicted Activity
4	1.086	1.142
10	0.537	0.904
12	0.721	0.981
13	0.148	0.266
18	0.886	1.007
22	1.366	1.086
23	0.958	1.272
28	1.221	0.878
34	1.455	0.949
35	1.468	1.130
39	0.920	0.931
50	1.200	1.112

Table 9. 3D QSAR derived predicted activities for c-Src.

Test molecules	Observed Activity	Predicted Activity
2	0.677	0.452
3	0.124	0.429
4	1.136	1.271
7	0.346	0.435
10	0.677	0.915
12	1.013	1.287
18	1.026	1.351
20	1.744	1.024
22	1.920	1.324
24	0.920	1.261
30	1.494	1.287
32	1.214	1.319
34	1.657	1.339
35	1.721	1.402
37	1.522	1.310
39	0.920	1.311
41	1.698	1.323

compounds belonging to the test sets of PDGFr, FGFr and c-Src cases, using Equations (5), (6) and (7), respectively.

The plot of observed versus predicted activities for the training and test sets of compounds in all the three cases are represented in Figures 5, 6 and 7. The external predictability of the QSAR models generated on test sets are characterized by pred_r^2 and pred_r^2se which are then calculated.

For PDGFr, pred_r^2 = 0.703, pred_r^2se = 0.159.

For FGFr, pred_r^2 = 0.553, pred_r^2se = 0.287.

For c-Src, pred_r^2 = 0.660, pred_r^2se = 0.352.

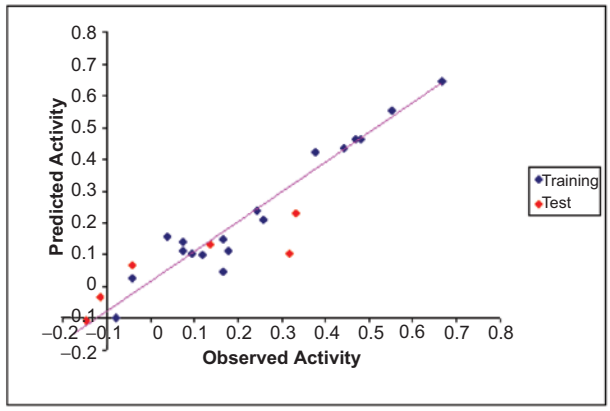


Figure 5. Plot of observed vs. predicted activities for training and test sets of PDGFr.

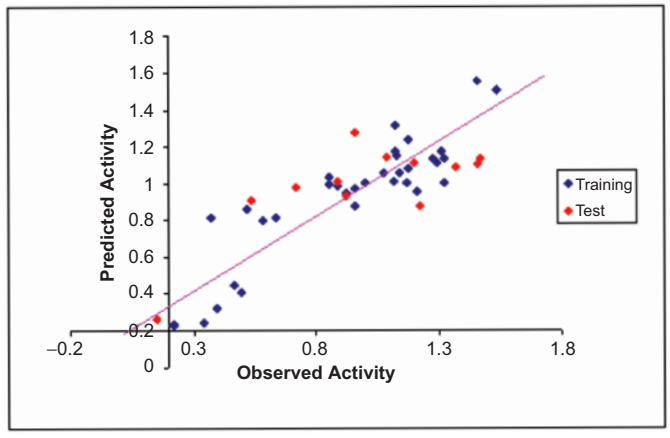


Figure 6. Plot of observed vs. predicted activities for training and test sets of FGFr.

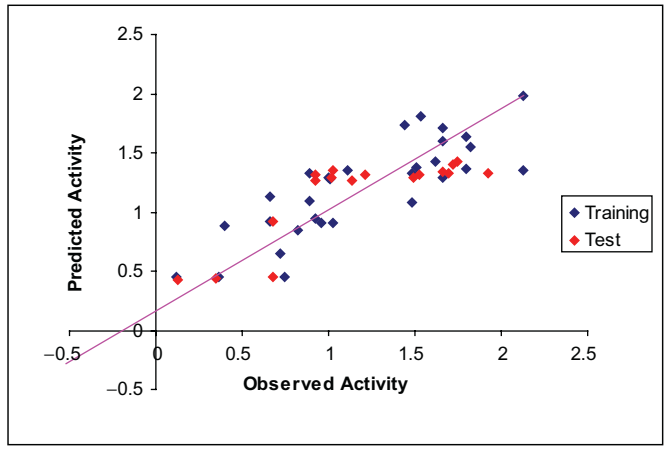


Figure 7. Plot of observed vs. predicted activities for training and test sets of c-Src.

The plot of observed versus predicted activities shows that the predicted activities of all the test compounds are in good agreement with their corresponding experimental activities and the fits are excellent.

Contribution plot of steric and electrostatic field interactions indicates relative regions of the local fields (steric and electrostatic) around the aligned molecules, leading to activity variation in the model [36]. The steric descriptor with positive or negative coefficients shows a region where bulky substituent is favored or disfavored, respectively. The electrostatic descriptor with a positive coefficient indicates a region favorable for electropositive group, while a negative coefficient indicates that an electronegative (electron donating) group is required at the position [37]. Figure 8 signifies contribution plot of steric and electrostatic field interactions for PDGFr.

From Equation (5) and Figure 8, it is clear that the presence of two steric descriptors (S_1368 and S_797) with positive coefficients indicates that the bulky substituents are favorable at C-6 and C-2 positions of the aligned molecules. The presence of electrostatic descriptors (E_1276 and E_1255) with positive coefficients and E_1516 with negative coefficient indicates electropositive and electronegative group should be substituted at the N-7 urea position. Figure 9 and 10 illustrate contribution plot of steric and electrostatic field interactions for FGFr and c-Src kinases.

Equation (6) and Figure 9 depict that electrostatic influences are most prominent in this case. Electrostatic descriptors with positive coefficients (E_802, E_972 and E_1812) indicate electropositive groups on C-2 and C-6 positions. N-7 urea substituent should possess electropositive as well as electronegative groups as indicated by the presence of electrostatic descriptors (E_2243 and E_1074) with positive coefficient and E_1097 with negative coefficient. Again, Equation (7) and Figure 10 illustrate that the steric effects are predominant in the third case where steric descriptors (S_1798, S_1952 and S_1112) with positive coefficients mean the presence of bulky substituents at C-2 and N-7 urea positions and the steric descriptors (S_604 and S_1808) with negative coefficients indicate that bulky groups are unfavorable at the 4th and 5th positions of the aminopyrido[2,3-d]pyrimidin-7-yl template.

Contribution plot of steric and electrostatic interactions elucidates the structure-activity relationship of these compounds. According to experimental findings, increased anticancer activity by the inhibition of PDGFr, FGFr and c-Src kinases were observed due to introduction of alkyl amino substituent at C-2, aryl substituent at C-6 and urea group at N-7 positions of the aminopyrido[2,3-d]pyrimidin-7-yl template. Compounds with long alkyl amino side chain at C-2, consisting of three carbons along with N-(4-methylpiperazino) as a terminator or four carbons terminating in N,N-dimethyl amino group moderately improves activity because the terminator possesses steric constraints around the distal amine [5]. It was also shown that crystal structures of these substituted urea derivatives maintain good activity due to forming intra molecular electrostatic field between the N-7 urea and N-8 of the pyrido pyrimidine template [5,38]. The experimental observations are strongly in lines with our theoretical findings based on contribution plots of steric and electrostatic field interactions shown in Figures 8, 9 and 10 where steric and electrostatic fields were enacting on C-2 and N-7 positions around the aligned molecules in both PDGFr and

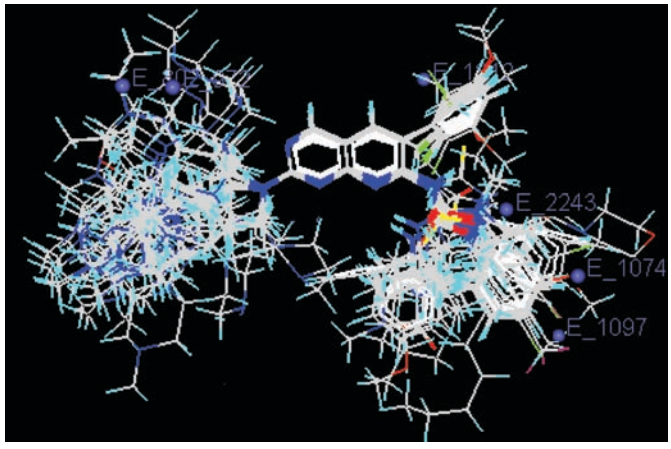


Figure 9. Contribution plot of steric and electrostatic field interactions for FGFr.

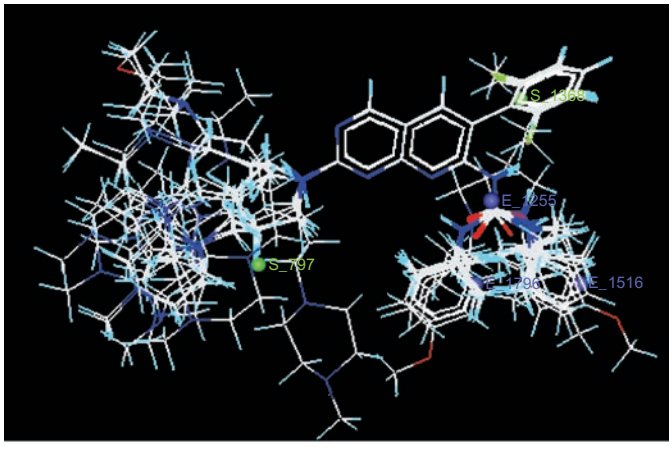


Figure 8. Contribution plot of steric and electrostatic field interactions for PDGFr.

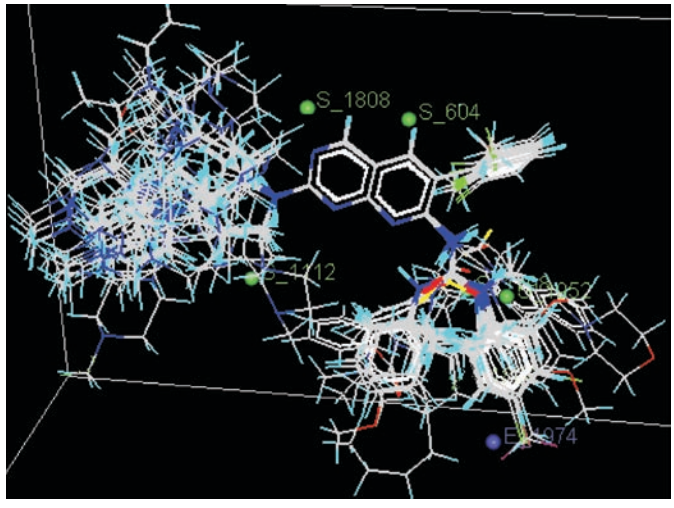


Figure 10. Contribution plot of steric and electrostatic field interactions for c-Src.

c-Src cases. Electrostatic field effects are most prominent in FGFr and steric influences are predominant in c-Src. This is also explained by 2D QSAR using ridge regression methodology where electrostatic field has significant R^2 value of 0.778 in FGFr but the electrostatic contribution in c-Src seems to be very low.

Conclusion

Ridge regression results show the contribution of different class of structural descriptors on anticancer activities of the aminopyrido[2,3-d]pyrimidin-7-yl derivatives. The 2D QSARs reported in Table 3 show that the calculated molecular descriptors can provide good quality predictive models for the compounds considered in the present investigation. Contribution plot of steric and electrostatic field descriptors generated by 3D QSARs shows the requirement of favourable group substituents in the 2-, 6- and 7- positions of aminopyrido[2,3-d] pyrimidin-7-yl template. In our effort to study the influence of steric and electrostatic descriptors in 3D QSAR models and contribution plots, it is observed that steric and electrostatic fields are acting on C-2 and N-7 positions around the aligned molecules of the PDGFr and c-Src kinases as shown in Figures 8 and 10. Again, for the FGFr and c-Src kinases, one is led to the fact that electrostatic fields are most prominent in FGFr and steric influences are predominant in c-Src. The present studies justify the development of models over the training set of compounds which are capable of predicting the activities of test compounds with reasonable accuracy. The successful development of predictive QSAR models in this direction affords rational anti cancer drug design.

***Declaration of interest*:** Sisir Nandi thanks the Council of Scientific and Industrial Research, New Delhi 110001, India for the grant of a Senior Research Fellowship.

References

1. Hamby JM, Connolly CJC, Schroeder MC, Winters RT, Showalter HDH, Panek RL, Major TC, Olsewski B, Ryan MJ, Dahring T, Lu GH, Keiser J, Amar A, Shen C, Kraker AJ, Slintak V, Nelson JM, Fry DW, Bradford L, Hallak H, Doherty AM. Structure-Activity Relationships for a Novel Series of Pyrido[2,3-d]pyrimidine Tyrosine Kinase Inhibitors. *J Med Chem* 1997; 40: 2296–2303.
2. Blume-Jensen P, Hunter T. Oncogenic kinase signaling. *Nature* 2001; 411: 355–365.
3. Klutchko SR, Hamby JM, Boschelli DH, Wu Z, Kraker AJ, Amar AM, Hartl BG, Shen C, Klohs WD, Steinkampf RW, Driscoll DL, Nelson JM, Elliott WL, Roberts BJ, Stoner CL, Vincent PW, Dykes DJ, Panek RL, Lu GH, Major TC, Dahring TK, Hallak H, Bradford LA, Showalter HDH, Doherty AM. 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.
4. Showalter HDH, Bridges AJ, Zhou H, Sercel AD, McMichael A, Fry DW. Tyrosine Kinase Inhibitors. 16. 6,5,6-Tricyclic Benzothieno[3,2-d] pyrimidines and Pyrimido[5,4-b]- and -[4,5-b]indoles as Potent Inhibitors of the Epidermal Growth Factor Receptor Tyrosine Kinase. *J Med Chem* 1999; 42: 5464–5474.
5. Schroeder MC, Hamby JM, Connolly CJC, Grohar PJ, Winters RT, Barvian MR, Moore CW, Boushelle SL, Crean SM, Kraker AJ, Driscoll DL, Vincent PW, Elliott WL, Lu GH, Batley BL, Dahring TK, Major TC, Panek RL, Doherty AM,Showalter HDH. Soluble 2-Substituted Aminopyrido[2,3-d] pyrimidin-7-yl Ureas. Structure-Activity Relationships against Selected Tyrosine Kinases and Exploration of in Vitro and in Vivo Anticancer Activity. *J Med Chem* 2001; 44: 1915–1926.
6. Bagchi MC, Maiti BC, Mills D, Basak SC. Usefulness of graphical invariants in quantitative structure–activity correlations of tuberculostatic drugs of the isonicotinic acid hydrazide type. *J Mol Model* 2004; 10: 102–111.
7. Bagchi MC, Maiti BC. On application of atom pairs on Drug Design. *J Mol Struct: THEOCHEM* 2003; 623: 31–37.
8. Bagchi MC, Maiti BC, Bose S. QSAR of antituberculosis drugs of INH type using graphical invariants. *J Mol Struct: THEOCHEM* 2004; 679: 179–186.
9. Nandi S, Bagchi MC. QSAR analysis of BABQ compounds via calculated molecular descriptors. *Med Chem Res* 2007; 15: 393–406.
10. Nandi S, Vracko M, Bagchi MC. Anticancer activity of selective phenolic compounds: QSAR studies using ridge regression and neural networks. *Chem Biol Drug Des* 2007; 70: 424–436.
11. Ghosh P, Thanadath M, Bagchi MC. On an aspect of calculated molecular descriptors in QSAR studies of quinolone antibacterials. *Molecular Diversity* 2006; 10: 415–427.
12. Hoerl AE, Kennard RW. Ridge regression biased estimation for nonorthogonal problems. *Technometrics* 1970; 8: 27–51.
13. Ajmani S, Jadhav K, Kulkarni SA. Three-Dimensional QSAR using the k-Nearest Neighbor method and its interpretation. *J Chem Inf Model* 2006; 46: 24–31.
14. Basak SC, Mills D, Hawkins DM, EI-Masri HA. Prediction of tissue:air partition coefficient: A comparison of structure-based and property-based methods. *SAR QSAR Environ Res* 2002; 13: 649–665.
15. Basak SC, Grunwald GD Niemi GJ. In: A.T. Balaban (Ed.). From Chemical Topology To Three-Dimensional Geometry. New York: Plenum Press; 1997. p 73–116.
16. http:// preadmet.brdrc.org/
17. Molecular Design Suite 3.5, VLife Technologies, Pune, India. www.vlifesciences.com
18. Roy K. Topological descriptors in drug design and modeling studies. *Molecular Diversity* 2004; 8: 321–323.
19. Katritzky AR, Petrukhin R, Tatham D, Basak S, Benfenati E, Karelson M, Maran U. Interpretation of quantitative structure–property and–activity relationships. *J Chem Inf Comput Sci* 2001; 41:679–685.
20. Rao CR. *Linear Statistical Inference and Its Applications* (2nd ed.). New York: Wiley;1973.
21. Randic M. Novel shape descriptors for molecular graphs. *J Chem Inf Comput Sci* 2001; 41: 607–613.
22. Rencher AC, Pun FC. Inflation of R2 in best subset regression. *Technometrics* 1980; 22: 49–53.
23. Estrada E. In: J Devillers and AT Balaban (Eds.). Topological Indices And Related Descriptors In QSAR And QSPR. Amsterdam, The Netherlands: Gordon and Breach; 1999. p 403–453.
24. Wold H. Soft modelling by latent variables: The non-linear iterative partial least squares approach. In: Gani J. (ed.). *Perspectives in probability and statistics. papers in honor of Bartlett MS*. London: Academic Press; 1975.
25. Hoskuldsson A. A Combined Theory for PCA and PLS. *J Chemom* 1995; 9: 91–123.
26. Massy WF. Principal Components Regression in Exploratory Statistical Research. *J. Amer Statist Assoc* 1965; 60: 234–246.
27. Frank IE, Friedman JH. A Statistical View of Some Chemometrics Regression Tools. *Technometrics* 1993; 35: 109–135.
28. Halgren, TA. Merck Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies. *J Comp Chem* 1996; 17: 553–586.
29. Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA) 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 1988; 110: 5959–5967.
30. Gasteiger J, Marsili M. Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges. *Tetrahedron* 1980; 36: 3219–3228.
31. Leach AR, Gillet VJ. An Introduction to Chemoinformatics., Boston: Kluwer Academic Publishers; 2003. p 80–100.
32. Leach AR. Molecular Modelling: Principles and Applications. Henry Ling, London; 2001. p 695.

33. NCSS – Statistical and Power Analysis Software; Hintze J. 2004 , NCSS and PASS. Number Cruncher Statitical Systems, Kaysville, Utah. http://www.ncss.com/
34. Bagchi MC, Mills D, Basak SC. Quantitative structure-activity relationship (QSAR) studies of quinolone antibacterials against M. fortuitum and M. smegmatis using theoretical molecular descriptors. *J Mol Model* 2007; 13: 111–120.
35. Hudson BD, Hyde RM, Rahr E, Wood J. Parameter based methods for compound selection from chemical databases. *Quant Struct Activity Relationships* 1996; 15: 285–289.
36. Kubinyi H. Ed. 3D QSAR in Drug Design. *Theory, Methods and Applications*. ESCOM, Leiden: Science Publishers B.V.; 1993. p 486–502.
37. Samee W, Nunthanavanit P, Ungwitayatorn J. 3D-QSAR Investigation of Synthetic Antioxidant Chromone Derivatives by Molecular Field Analysis. *Int J Mol Sci* 2008; 9: 235–246.
38. Trumpp-Kallmeyer S, Rubin JR, Humblet C, Hamby JM, Showalter HDH. Development of a Binding Model to Protein Tyrosine Kinases for Substituted Pyrido[2,3-d]pyrimidine Inhibitors. *J Med Chem* 1998; 41: 1752–1763.