Combined Ligand and Structure Based Approaches for Narrowing on the Essential Physicochemical Characteristics for CDK4 Inhibition

Nahren Manuel Mascarenhas and Nanda Ghoshal*

Structural Biology and Bioinformatics Division, Indian Institute of Chemical Biology (CSIR), 4 Raja S.C. Mullick Road, Jadavpur, Kolkata - 700032, India

Received January 28, 2008

In the absence of an experimentally determined 3D structure of CDK4 (Cyclin-Dependent Kinase 4), QSARs (Quantitative Structure Activity Relationship) have been explored to rationalize binding affinity in terms of physicochemical and structural parameters. Further, docking on a homology model of CDK4 validated the derived QSARs and predicted the binding mode of this series of inhibitors. Relevant parameters and leave-one-out (LOO) cross-validation (q^2) as well as an external test set validation (r^2_{pred}) judged the statistical significance and predictive ability of the models. Docking enabled a better understanding of protein−ligand interaction and provided a mechanistic interpretation in terms of physicochemical characteristics. It identified a unique hydrogen bonding between the imidazole of His-95 and the pyridine nitrogen in the ligand. It rationalized the need for R_2 substituents to be bulky and polar, while the substituent at R_8 to be hydrophobic and comparatively less steric. It also explained why at R_6 a variety of substituents are tolerated and how the presence of methyl at R_5 enhances binding affinity.

INTRODUCTION

The cell cycle events are well coordinated resulting in proper DNA replication, leading to healthy daughter cells. The cell cycle is a frequent target of genetic alterations in cancer because of its central role in the control of cell growth and proliferation. The central players in the cell cycle machinery are Cyclin-Dependent Kinases (CDKs). CDKs are a family of heterodimeric serine/threonine kinases consisting of a catalytic CDK and an activating cyclin subunit. In mammals, there are different CDKs controlling different stages of the cell cycle and play the role of switch on/off on the cell cycle process. When on, the cell passes through the stage that the particular CDK controls, while when off, the cell cycle stops when it reaches the stage controlled by that CDK.[1]

Several identified CDKs are known to intervene in cell cycle machinery at different phases of cell cycle.[2] One of the most important CDK substrate in mammalian cells during the G_1 phase is the product of retinoblastoma tumor suppressor gene (pRb). The RB family of proteins (pRB, p107, and p130) acts as brakes on the cell cycle machinery and is capable of restricting the cell cycle at G_1 phase.[3] At the beginning of G_1, pRb proteins exist in an active, nonphosphorylated state that is able to bind to a large number of proteins. During the G_1 phase, pRb proteins are sequentially phosphorylated, mediated initially by cyclin-D dependent CDK4/6 kinases. Phosphorylation of pRb leads to their partial inactivation, which allows transcription of E2F-controlled genes. CDK4 and CDK6 are amplified or overexpressed in a variety of tumors and are frequent targets of genetic alterations in cancer. In addition, the natural inhibitors of CDK4/6, INK4a-ARF, and INK4b loci are altered in a variety of human cancers. Also, overexpression of D-type cyclins has been observed in different types of neoplasias such as breast tumors, esophageal cancer, colorectal carcinomas, head and neck carcinomas, and lung cell tumors.[4,5] Further, CDK4/6-CyclinD complexes, very recently, have been demonstrated as bonafide cancer targets, especially for breast cancer. [6,7]

Understanding the influence of chemical structure on biological activity is of paramount importance. For this reason, quantitative structure activity relationship (QSAR) is considered a major tool in drug discovery. It represents a technique to understand the relationships between chemical structures and their biological property by means of descriptors calculated for the molecules of interest and enables the modification of the lead structure to design compounds to enhance potency. The results, also, will help in predicting in silico the biological activity of newly designed molecules. Deducing QSAR becomes more significant when no structural information is available for the target, as is the present case. Though experimental structure information is available for CDK2 and CDK6, till date no 3D structural information is available for CDK4. In such a case, ligand based drug design approaches prove to be versatile tools.

The major objectives of the study were (i) to enable a better perspective of the various physicochemical aspects of protein−ligand interactions that exist in this system and (ii) to formulate practically useful, easily interpretable QSAR models that can guide chemist in designing better inhibitors. Since these structures present a common core, a fragment based QSAR approach has been employed. In addition, due to the high sequence similarity shared between CDK6, CDK2, and CDK4, a protein model of CDK4 has been constructed based on homology modeling. With the aid of docking, we have studied the mode of binding as well as the interactions that ligands compose with the protein and

* Corresponding author phone: +91332473 3491 (extn 254/236); fax: +91332473 5197; e-mail: nghoshal@iicb.res.in.

10.1021/ci8000343 CCC: $40.75
Published on Web 06/20/2008

have correlated the conclusions drawn from docking to those established in QSAR.

MATERIALS AND METHODS

Data Set. The data set, constituting the basis for QSAR model generation, consisted of 129 derivatives of pyrido[2,3-d]pyrimidin-7-ones, compiled from three publications reported in the literature with biological testing performed using the same assay,[8–10] and their molecular structures are provided in Table 1 (in the Supporting Information). The compounds spanned an activity range of nearly 4 log units making them constructive for QSAR studies. The IC_{50} values of all molecules were converted to pIC_{50} [=log(1/IC_{50})] that served as the dependent variable.

Molecular Modeling. All molecules were modeled in Catalyst (v 4.11),[11] running on SGI fuel workstation with IRIX 6.5 OS. The molecules were converted to optimal 3D-structure using CORINA,[12] implemented in TSAR (v 3.3) software. [11]

Division of Data Set: Clustering. Clustering guides in classifying a collection of heterogeneous data into groups, such that each group is a homogeneous population of subdata of similar nature. Clustering, performed on molecules considered for the study, guided in dividing the data set into training and test sets. The rationale for adopting such a protocol was to ensure both sets cover the entire multidimensional descriptor space. Whole molecule descriptors, used for clustering, were calculated using MOE[13] and QSAR+ module of Cerius[2].[11] One hundred descriptors, which included topological, structural, information, and group counts, were calculated. For Principal Component Analysis (PCA) descriptor values were scaled to a variance of 1.0. Three principal components were derived from PCA and were used to cluster the complete data set by the hierarchical cluster analysis[14] (HCA-Ward) method using the Cerius[2] (v 4.10) software. Based on clustering methodology, we intended to set aside one-fifth of the total data set as the test set. The test and training sets were selected in such a way that all clusters were represented in both sets. From each cluster that had more than three elements, the test and training sets were divided such that both cover the entire range of activity encapsulated in that cluster. This led to 108 molecules in the training set with the remaining 21 making up the test set [Figure 1].

QSAR Methods. *Definition of Fragments.* Two issues need attention if one is interested in generating a clear SAR interpretation.[15] First is the knowledge of what features of structures are measured by a given descriptor. The second is the knowledge of where the changes in structure are taking place that can be correlated to changes in the experimentally observed property (pIC_{50} in this case). In case of whole-molecule descriptors, changes in the features of structures measured by the descriptor may be occurring in several places. Hence changes in descriptor values need not directly reflect the changes accompanying at a particular substructural level although the changes affecting the observed property may be localized in a particular region of the molecule. In this context, fragment based QSAR models are of potential value, as it would help in identifying clearly the biological effect of substructural changes accompanying the molecule in an intuitive manner, and, therefore, the challenge lies in the way fragments are defined. The present collection of

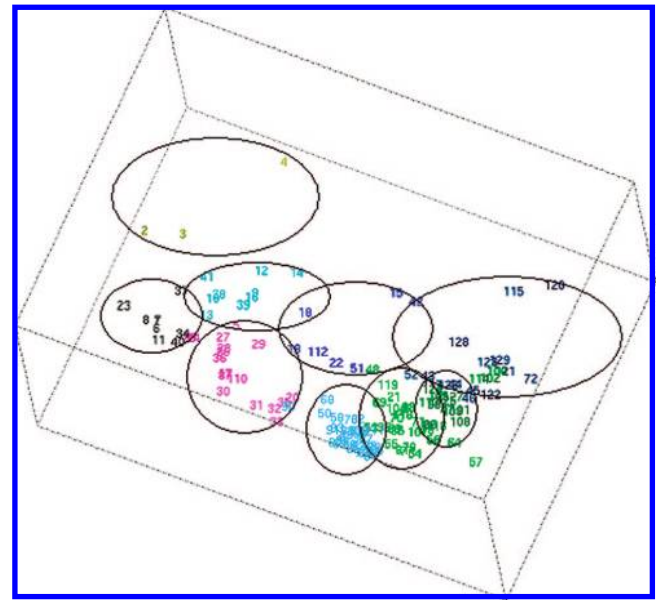


Figure 1. Pictorial representation of clustering result. Compounds belonging to each of the clusters are shown in nine unique colors.

molecules revolves around pyrido[2,3-d]pyrimidin-7-one as the core, and four fragments R_2, R_5, R_6, and R_8 were defined as depicted in Table 1. All descriptors included in the study were calculated for each of these four fragments for each molecule.

Descriptors Selected for the Study. Although there exist many descriptors, which vary in complexity, the information they encode, and the time required in calculating them, only a few could relate to the physicochemical attribute of molecules. The foremost purpose of devising a QSAR, especially in the absence of any structural information for the target, is to understand the physicochemical role of molecules against the measured activity/property. This directly depends upon the nature of descriptors included for characterizing the attribute of the fragment as that dictates the chemical insight to be gained from the QSAR. Combining the fragment based approach with physicochemical descriptors one should, in principle, be able to conclude upon the true nature of protein−ligand interactions at that particular substructural region. This would in turn guide in understanding the physicochemical contribution of each fragment and would aid in modifying or designing new lead molecules.

Table 2 lists the descriptors selected for the study. The logarithm of partition coefficient function (logP)[16] has been known to have a direct relationship with ligand binding and is one of the factors in determining the ADME (Absorption Distribution Metabolism Excretion) properties of the molecules. The molar refractivity describes the steric nature of the molecule or substituent. Sterimol parameters, proposed by Verloop[17] are a set of multidimensional steric parameters of the substituent. It assumes that all atoms have van der Waals radii and uses these to define the substituent's space requirements. Sterimol parameter L, known as the length parameter, defines the maximum length of the substituent along the axis of the bond between the first atom of the substituent and the parent molecule. The other parameters, represented by VB(1−5), are known as the width parameter. Hydrogen-bonding interactions are a very important part of ligand−receptor interactions, so the number of hydrogen-

Table 2. Descriptors and Indicator Variables Used in the Study

logP	calculated octanol/water partition coefficient	
MR	molar refractivity represents steric properties	
MM	molecular mass of the fragment	
MV	molecular volume of the fragment	
NHBA	number of hydrogen bond acceptor atoms	
NHBD	number of hydrogen bond donor atoms	
Verloop descriptors		
L(VL)	describes the length parameter of the substituent	
B(VB1-VB5)	describes the width parameter of the substituent	
bond lipole	is measure of lipophilic distribution of the substituent	
I_p	1 for the presence of positive ionizable group at R_2	and 0 otherwise
I_2	1 for the presence of pyridine ring at R_2	
I_5	1 for the presence of methyl fragment at R_5	
I_6	1 for the presence of any substituent at R_6	

Figure 2. The figure represents alignment of the template (CDK2 and CDK6) with target (CDK4). The figure also depicts the conserved regions within CDK6 and CDK4. The hinge region constituting the active site is boxed.

bond acceptors (NHBA) and donors (NHBD) were also included as descriptors. Another descriptor that was considered for deriving QSAR was bond lipole (BL). BL is a measure of the lipophilic distribution of the substituent calculated from the summed atomic logP values. Other descriptors considered for deriving QSAR include molecular mass (MM) and molecular volume (MV). All fragment based descriptors considered in the present QSAR study were calculated using TSAR software. Ionizable molecules were taken in their neutral form due to difficulty in calculating the descriptor values for ionizable forms.

Indicator Variables. Indicator variables are descriptors that can be assigned only two values, 1 and 0. They take a value of 1 for the presence and 0 for the absence of a particular moiety and help in explaining the importance of that moiety toward biological activity. For compounds containing a positively ionizable group, such as piperazine at the R_2 position, the calculated pK_a[18] values were in the range of 8.5−9.5. This implied that these fragments will be positively charged at biological pH. In such cases the contribution of this positive ionizable group remains unaccounted, and, for this reason, indicator variable I_p was defined. The indicator variable I_2 was defined to highlight the role of the pyridine nitrogen toward ligand binding. The original authors[9,10] observed only methyl to be optimal at R_5, while a variety of substituents were found to be tolerant at R_6 irrespective of their physicochemical nature. Hence, two variables, I_5 and I_6, were defined to imply the importance of substitution at R_5 and R_6, respectively.

QSAR Model Construction. Stepwise Multiple Linear Regression (SMLR) and Genetic Function Approximation[19] (GFA) were utilized as statistical tools to arrive at the most appropriate set of descriptors that are able to explain the variance in biological activity. SMLR analyses were performed using SPSS,[20] while GFA was operated as implemented in the QSAR+ module of Cerius[2], which also allows the use of spline terms for nonlinear modeling. For SMLR, the F to include was set to 4, while the F to exclude was set at 3.5. The GFA settings, number of crossovers = 25000, linear/spline terms, and smoothness factor (d) = 1.0 were used. The spline terms, denoted with angle brackets in the QSAR equation, take the form, <a-f(x)>, where "f(x)" is the descriptor used as the independent variable, while "a" is referred as the knot of the spline. In the present case, 300 equations were generated in a single GFA run, and the equation with the lowest lack of fit (LOF) value was considered for further discussion.

Homology Modeling. CDK6 and CDK2 share good homology with CDK4, about 68% and 46%, respectively [Figure 2]; therefore, it was sensible to construct a homology model based on these two structures. The cyclin subunits were truncated from both CDK6 (PDB ID: 2EUF)[21] and CDK2 (PDB ID: 1FIN),[22] and using the Modeler program implemented in InsightII 11, ten homology models were generated. The best model determined with the lowest value of the Modeler objective function was minimized with the CVFF (Consistent Valence Force Field) by the steepest descent algorithm method until an energy gradient of rms

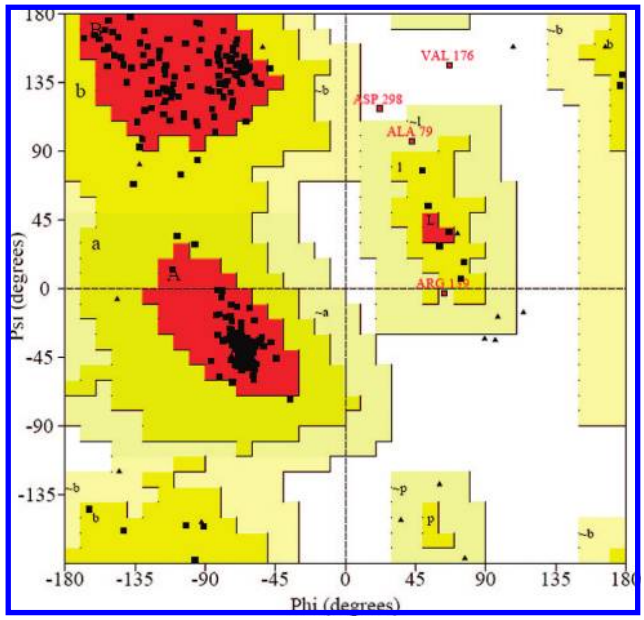


Figure 3. Phi-Psi plots of the modeled CDK4 obtained by Procheck. Two residues in the generously allowed regions Arg159 and Ala79 as well as in the disallowed regions Asp298 and Val176 are labeled.

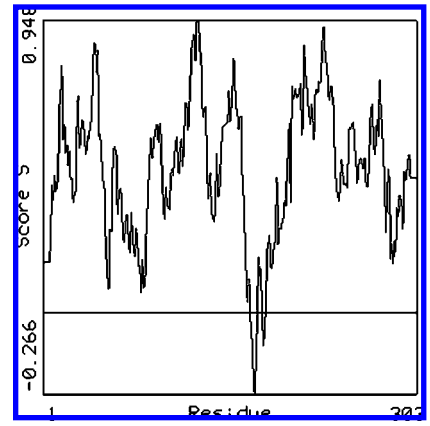


Figure 4. Graph showing the result of profile-3D calculation of the modeled CDK4.

(Root Mean Square) −0.100 was obtained, and the subsequent model was used for docking. The resulting model had 90.1% of the residues within the most favored region of the Ramachandran plot [Figure 3].[23] Checking by profile-3D showed that the overall self-compatibility score for this protein was 128.33 which was much higher than the minimum score of 62.05 and close to the expected score of 137.89 [Figure 4]. It is to be noted that a compatibility score above zero indicates the presence of residues in the favorable regions. Nine residues namely Met169-Val174 and Leu178-Tyr180 were below zero, but, since these lie far away from the active site of CDK4, they will practically have no impact on this study.

Docking. GOLD (Genetic Optimization for Ligand Docking)[24] was used for all our docking experiments. For each ligand, 50000 genetic operations were performed on a population size of 100 individuals with other default parameters (selection pressure of 1.1 and crossover, mutation, and migration as 95, 95, and 10, respectively). The binding site was defined as all protein atoms which are 10 Å away from carbonyl oxygen of Val-96. The final conformation of the ligand was selected first, considering the conformation with the highest scoring function and then considering the one that possesses hydrogen bonding interaction with hinge region. The Chemscore fitness function implemented in GOLD was used, while Ludi3[25] scores were calculated using Discovery Studio (v1.7). The selected poses were then minimized within the ATP pocket considering protein residues that lie within 8 Å from any ligand atom as flexible. The minimization using CVFF was performed with the steepest descent method until a gradient of 0.1 was achieved. The resulting protein−ligand conformations were used to calculate the intermolecular interaction energy.

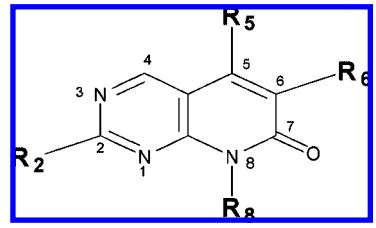
RESULTS AND DISCUSSION

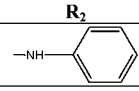
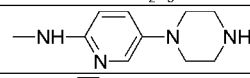
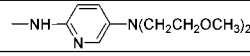
The objective of the present study was to rationalize binding affinity data of CDK4 inhibition into meaningful interpretable models with the help of QSARs. To derive highly statistically significant models with completely uninterpretable correlations was not our aim. We feel a QSAR study failing to address the molecular basis of action for what it is derived is not practically useful. We, therefore considered such a type of descriptors and statistical tools to make our result more appealing to a medicinal chemist looking to derive a good understanding of biological phenomena from QSARs. Table 3 represents a few examples to demonstrate the calculated descriptors and the result of QSAR models. As demonstrated, the fragments are truly described according to their physicochemical characteristics. Such an effective description of molecules into interpretable descriptors should elate the understanding of this inhibitory mechanism.

Table 4 lists the five QSAR equations generated in the study. The predicted pIC_{50} of the training and test set compounds of the five models are tabulated in Table 5 (in the Supporting Information). Several statistical parameters considered to ensure the models were significant (Table 6). These include (i) correlation coefficient (r^2), (ii) explained variance (r_a^2), (iii) standard error of estimate (s), and (iv) variance ratio (F). Although many statistical parameters exist to check the consistency and the robustness of the model, a true guide to evaluate the stability and internal predictivity of the model is (i) cross-validation r^2 (q^2) performed by the leave one out (LOO) approach, (ii) predicted residual sum of squares (PRESS), and (iii) standard error of prediction (s_{PRESS}). The widely accepted parameter to reflect the true correlation is data randomization also known as *y-scrambling*. In this technique, the existing values of dependent variable are shuffled within the data set, and the models are regenerated with this scrambled data set. The result, if worse than the original unscrambled data set, truly reflects the absence of chance correlation, else it strongly suggests a significant correlation by chance. In each of the generated models, care was taken to exclude the appearance of correlated descriptors within the same equation (Table 7 in the Supporting Information).

Another parameter usually calculated to address the predictive confidence of the models is predictive r^2 (r^2_{pred}). The r^2_{pred} was calculated for each molecule in the test set using the equation $r^2_{pred} = 1 - (PRESS/SD)$ where PRESS is the sum of the squares of the residuals of every compound in the test set and SD is the sum of the squares of the difference between the experimental value of the test set and

Table 3. Few Structures as Examples, Showing the Calculated Fragment Descriptors and Their Prediction with QSAR Models



ID	pIC_{50}	Fragment R_2	Descriptor $MR(R_2)$	Descriptor $LogP(R_2)$	Prediction Model 1	Prediction Model 4	Prediction Model 5
1	6.21		28.774	1.557	5.95	5.68	5.89
2	4.92	$-NH-C_2H_5$	13.745	0.218	5.36	5.03	4.81
113	7.96		51.394	0.175	7.61	7.57	7.71
120	7.29		62.726	0.406	7.29	7.30	7.11

ID	pIC_{50}	Fragment R_6	Descriptor $LogP(R_6)$	Descriptor $BL(R_6)$	Descriptor $VB1(R_6)$	Prediction Model 1	Prediction Model 3
89	7.80	-Br	1.236	2.338	1.900	7.32	7.52
93	7.29	-F	0.584	0.789	1.550	7.55	7.57
102	7.43	$-O(CH_2)_2OEt$	0.369	2.149	1.500	7.62	7.15
103	5.75	$-OCH_2{}^iPr$	1.406	5.433	1.785	7.27	6.58
104	6.91	-Ac	-0.510	-0.881	1.650	7.93	8.12
105	6.23	-COOEt	2.537	1.499	1.728	7.66	7.57

ID	pIC_{50}	Fragment R_8	Descriptor $MR(R_8)$	Descriptor $LogP(R_8)$	Descriptor $MM(R_8)$	Prediction Model 1	Prediction Model 4
1	6.21	-Et	10.445	0.263	29.07	5.95	5.68
30	6.68	-cyclopentyl	22.109	1.109	69.14	6.34	6.83
31	7.33	-cyclohexyl	26.710	1.506	83.17	6.52	6.59
34	5.78	-phenyl	25.474	1.602	77.11	6.56	6.58
35	4.87	-cyclohexylmethyl	31.441	1.819	97.20	6.66	6.33
36	6.03	-benzyl	30.309	1.697	91.14	6.61	6.31
37	5.38	-methoxymethyl	11.550	0.123	45.07	5.29	4.87

Table 4. Five QSAR Equations Developed in our Study

model	QSAR equations	no. of free variables
1 (SMLR)	$pIC_{50} = 4.799 + 0.786I_6 + 0.397I_p - 0.023MM(R_5) - 0.345logP(R_6) + 0.460logP(R_8) + 0.039MR(R_2) - 0.362NHBA(R_2) - 0.598NHBA(R_8)$	28
2 (SMLR)	$pIC_{50} = -0.892 + 0.699I_6 + 0.331\ I_P - 0.034MM(R_5) - 0.083BL(R_2) - 0.210BL(R_6) + 0.046MR(R_2) + 3.609VB1(R_8) - 0.522NHBA(R_2) - 0.742NHBA(R_8)$	28
3 (GFA)	$pIC_{50} = -1.892 + 0.046MR(R_2) + 0.356I_p - 0.264BL(R_6) + 1.040VB1(R_6) - 0.747NHBA(R_8) - 0.526NHBA(R_2) + 3.599VB1(R_8) - 0.084BL(R_2) - 0.036MM(R_5)$	60
$4(GFA_{spline})$	$pIC_{50} = 6.063 - 1.267NHBA(R_8) + 1.057 < 0.210 - logP(R_6) > - 0.037MM(R_5) - 0.472 < NHBA(R_2) - 1 > - 0.043 < 56.259 - MR(R_2) > + 0.029MM(R_8) - 0.140 < MR(R_8) - 22.109 >$	28
$5(GFA_{spline})$	$pIC_{50} = 8.744 - 0.601NHBA(R_2) - 0.644\ NHBA(R_8) - 0.032MM(R_5)\ 0.432 < 9.356 - VB5(R_2) > - 7.297 < 1.811 - VB1(R_8) > + 0.900 < 0.210 - logP(R_6) > - 1.636 < logP(R_2) - 1.696 >$	60

the mean of the training set. In our case 21 molecules made up the test set, and the results are represented in Table 6.

Model 1(SMLR). The first model was derived using some of the simple descriptors commonly used in QSAR viz., logP, MR, MM, NHBA, NHBD, and indicator variables. The model had a goodness of fit value of 0.607, could explain 57.5% of variance, and possessed a q^2 value of 0.527 [Figure 5] indicating the model to be statistically reliable. In addition, the standard error of estimate (s) of 0.654 and the standard error of prediction (s_{PRESS}) of 0.718 further signify the statistical confidence. LogP descriptor is a measure of hydrophobicity and could be correlated to a lipophilic type of interaction with macromolecules. One could correlate the appearance of logP (R_8), bearing a positive coefficient, to the existence of such type of interactions between the ligand and protein at this region. The presence of logP (R_8) and NHBA (R_8), bearing a positive and a negative coefficient, respectively, converge on the point that a hydrophobic

Table 6. Statistical Parameters of the Models

model	r^2	LOF	r_a^2	q^2	s	F(df)	PRESS	s_{PRESS}	randomi-zation (r^2)	r^2_{pred}
1 (SMLR)	0.607	-	0.575	0.527	0.654	19.113(8,99)	50.492	0.718	0.069	0.627
2 (SMLR)	0.712	-	0.685	0.641	0.563	26.892(9,98)	38.310	0.628	0.091	0.640
3 (GFA)	0.718	0.401	0.692	0.648	0.557	27.762(9,98)	37.521	0.622	0.074	0.653
4(GFA_{spline})	0.675	0.463	0.652	0.624	0.592	29.608(7,100)	40.074	0.636	0.231	0.700
5(GFA_{spline})	0.734	0.378	0.716	0.695	0.563	39.502(7,100)	32.580	0.549	0.238	0.641

fragment is more favored at this region. Consequently, fragments that possess polar atoms as in compounds 37−41 seem to be unfavorable and propose the presence of a hydrophobic interaction to contribute to ligand binding. Although the appearance of I_6 in the equation suggests the mere presence of a substitution at R_6 to be constructive for biological activity, it fails to speculate about the type of fragment preferred at this region. The descriptor log (R_6) which appears in the equation with a negative coefficient describes the physicochemical requirement at this region. This suggests that fragments whose logP values are above zero would contribute negatively, while polar fragments whose logP values fall below zero will contribute positively to the QSAR equation. Hence, one could contemplate the presence of some polar type of interactions at this region to stabilize the protein−ligand complex.

Model 2 (SMLR). In an attempt to improve the equation statistics and predictive ability of the model, Verloop's STERIMOL parameters and bond lipole descriptors were included during the SMLR run. From the statistical viewpoint, Model 2 is appreciably better than Model 1 as reflected in the value of r_a^2 (68.5%), s (0.563), F (26.892), and q^2 (0.641). Two new descriptors that appear in this equation are bond lipole descriptors, BL (R_2) and BL (R_6). Both BL (R_2) and BL (R_6) appear in this equation with a negative coefficient suggesting that the lipophilic substituents at these positions are unfavorable toward biological activity. Although Model 1 testifies hydrophobic substituents at R_8, it fails to shed light on the steric requirement at this region. The appearance of VB1 (R_8) with a positive coefficient addresses this issue and suggests that steric groups are tolerated at this region. A clear picture about R_8 can be drawn by combining Models 1 and 2, a preference for hydrophobic substituents and the tolerance toward steric moieties.

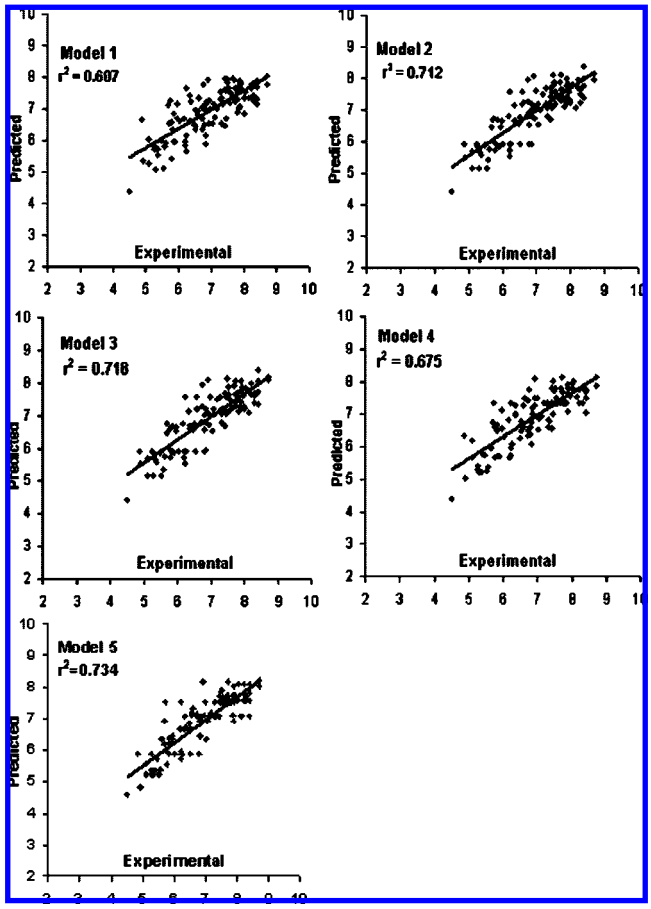


Figure 5. The plots of experimental (pIC_{50}) vs predicted (pIC_{50}) for the five QSAR models derived in the study.

Model 3 (GFA). This QSAR model derived by virtue of GFA appears to follow a similar pursuit to Model 2. The only exception visible in this model is the appearance of VB1 (R_6) in place of I_6. Although the mere presence of I_6 in the QSAR equations could be related to the necessity of substitution at R_6, it fails to define the nature of substituent preferable at this position. As Verloop's STERIMOL parameter (VB) reflects the width of the substituent, the presence of VB1 (R_6) with a positive coefficient suggests that the substitution at this region should be bulky in nature for optimal inhibition. The tolerance of bulky substituents, as proposed by these models, indicates the presence of a cavity at this region of the protein. These results when coupled to that obtained from model 1 propose that hydrophobic groups with polar atoms may have an advantage over hydrophobic moieties. The exact nature of interactions and their role in enhancing protein−ligand affinity is addressed in the docking section of the manuscript.

Four descriptors, I_p, MM (R_5), NHBA (R_2), and NHBA (R_8), appear in all the QSAR equations discussed so far. The indicator variable I_p appears in these equations with a positive coefficient; this implies the favorable role of a positive ionizable group at this region. From the perspective of a receptor, one could predict a negatively charged amino acid to interact with this positively ionizable group that can result in tight binding arising from strong Columbic interactions. In the present study this fragment, in most cases being the piperazine nitrogen, can attain a positive charge at biological pH with its pK_a value around 9.0. MM (R_5) appears in all these equations with a negative coefficient suggesting substitution at R_5 to be detrimental for optimal activity. The presence of heavy substituents, as in 66 and 67, decreases the potency nearly 50- and 200-fold, respectively, indicating the binding pocket at this region to be small and intolerant to further substitution[9] (refer to Model 6 in the Supporting Information). The presence of NHBA (R_2) and NHBA (R_8) in these equations with a negative coefficient specifies that hydrogen bonding acceptor groups are unfavorable for biological activity. The presence of NHBA (R_8) with negative coefficient substantiates the earlier argument that the pre-

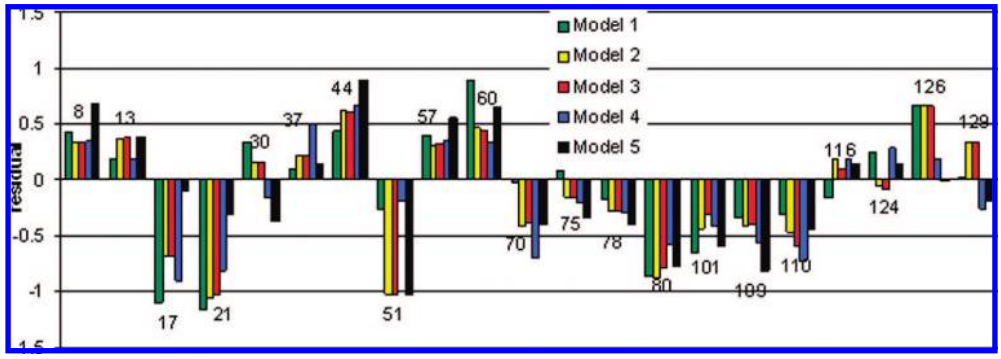


Figure 6. Plotted are the residual values of the test set for the five models.

ferred fragment at R_8 should be hydrophobic. Based on these conclusions with respect to the R_8 region, one can sense the possibility of a hydrophobic interaction to contribute to overall binding.

Model 4 (GFA-Spline). A nonlinear model (4) was generated with the aid of 'splines' to circumvent the drawbacks of linear modeling and with the intention to explore nonlinear relationships. For this case, the same set of descriptors used to derive Model 1 was considered. This model proved superior to Model 1 in terms of statistics and predictive ability. The model could now explain 65.2% of variance and possessed a better q^2 value of 0.624. Compared to Model 1, the s and s_{PRESS} values decreased to 0.592 and 0.566, respectively, while the F and r^2 values increased. The model possessed a better predictive ability, reflected in the value of predictive r^2 (r^2_{pred}=0.700) [Figure 6]. The model has also offered some interesting physicochemical meaning for a better understanding of protein−ligand interactions that remained undisclosed in the linear modeling approaches.

Prior to further discussion of the model, it is essential to understand the difference between the two spline terms, <a-f(x)>and <f(x)-a>. In the first case, the spline term will be zero as long as "a" is ≤ "f(x)", while in the second case, the spline contributes zero till "f(x)" ≤ "a". In other cases, the spline will take the value of (a-f(x)) or (f(x)-a). The appearance of logP(R_6) within a spline term with a knot value of 0.210 specifies that if a fragment's logP value is more than 0.210, the contribution of this term will be negated. Hence fragments with a logP value less than 0.210 contribute to activity. Another important outcome from this spline is that if fragments possess a hydrophobic moiety, it will in principle not contribute negatively toward binding. Hence, the equation hypothesizes although polar fragments are optimal, hydrophobic fragments are tolerated without significant loss in activity at R_6. This is in clear contrast to previous equations, which suggest a hydrophobic fragment to be detrimental. Hence, the use of splines appears to describe a better interpretation of the underlying SAR. NHBA (R_2) appears in the equation as a spline term, with a knot value of 1 and with a negative coefficient, and advocates that when R_2 possesses 1 or 0 hydrogen bond acceptors, the contribution of the spline term will be zero. But when NHBA (R_2) becomes greater than 1, the spline will contribute negatively toward CDK4 inhibition. So the preferred fragment at this position should not be a substituent with more than one hydrogen bond acceptor group. MR (R_2) appears as a spline with a knot value of 56.259. The spline term gives a clear picture on the importance of bulky groups at this position. It suggests that the smaller the fragment is at this position, the more negatively it contributes to biological activity. The contribution of substituent R_8 is reflected in terms of MM (R_8) and MR (R_8). Since MM (R_8) appears in the equation with a positive coefficient, it suggests that a bulky substituent may be acceptable at this position. But this variable does not reflect the critical bulkiness that may be tolerated and is indicated clearly by MR (R_8) that appears as a spline term in the equation. This spline term reveals that as long as the MR (R_8) value is less than 22.109 the spline will contribute zero toward biological activity, while a value of MR (R_8) above the knot value will contribute negatively towards binding. A comparison drawn between 34, 35, 36, and 37 enables a better understanding of the optimal requirement at R_8. These possess phenyl, cyclohexylmethyl, toluene, and methoxymethyl fragments at R_8, and correspondingly their MR values vary as 25.47, 31.44, 30.31, and 16.74, respectively. Except 37, other molecules possess substituents whose MR value exceed the critical value of 22.109 and are predicted rightly as inactive. In the case of 37, although the MR value falls within the critical bulkiness that could be accommodated at this region, the presence of one NHBA in methoxymethyl makes it unfavorable at this region since NHBA appears with a negative coefficient in this equation.

Model 5. Encouraged by the result obtained by inclusion of splines, the following equation (5) was obtained by including all descriptors considered for the study. The model could now explain 71.6% of variance in the dependent variable. A low value of s_{PRESS} (0.549) coupled with a high value of q^2 (0.695) and r^2_{pred} of 0.641 authenticates the true predictive nature of this model.

In this model, VB5 (R_2) appears with a knot value of 9.356, and, since VB5 points to the maximum width of the substituent, the presence of the bulky substituent at R_2 appears to be critical for optimal activity. This spline term also highlights that if the VB5 (width) value of the substituent is above the value of 9.356, it will be favorable. Viewing from another perspective this equation also suggests that the lower the value of VB5 (R_2) the more unfavorable it will be toward ligand binding. Similarly, the presence of VB1(R_8) with a knot value of 1.811 suggests that there is room to accommodate substituents with the VB1(R_8) value above 1.811 at the R_8 region. VB1 is always less than VB5 (except in the case of a single atom substituent) as it is a measure of the smallest width of the substituent in any direction. Hence this result also corroborates with Model 4 with regard to the difference in the steric tolerance at R_2 and R_8. Log(R_2) appears in this equation with a negative coefficient along with a knot value of 1.696, suggesting substituents with a

Table 8. Outliers of Five QSAR Models

model	outlier compounds
1	35,47,103,105,128
2	32,43,84,103,104, 105,128
3	32,43,103,104,105,119,128
4	35,45,69,103,104
5	35,43,45,103,104,105,119

logP value greater than 1.696 contribute negatively toward CDK4 inhibition. NHBA (R_2) appears with a negative coefficient, suggesting the presence of hydrogen bond acceptor atoms at this region to be detrimental to biological activity.

Common Outliers. Careful inspection of the outliers is necessary to gain a better appreciation of the derived QSARs and more importantly to understand its limitations.[26] The most common explanation proposed for the outcome of outliers in QSAR is based on the finding that these outliers bind to the receptor in a different binding mode and hence do not follow the trend pursued by other molecules included in the study.[27] Other reasons proposed for their occurrences in QSAR studies are an incorrectly measured experimental value, a significant difference in the physicochemical property, and their structural uniqueness. Outliers may even arise due to the lack of certain descriptors or parameters included in the study that are unable to describe the QSAR for the compounds in the entire data set. Outliers may also be due to the inadequacy of the modeling technique used to derive the model. In the current study few outliers appeared in the model development process as tabulated in Table 8.

Compounds 32, 35, 43, 45, 103, 104, 105, 119, and 128 appeared as outliers in any two of the five models and were termed as common outliers. Note that compounds 47 and 69 also appeared as outliers but only in models 1 and 4, respectively. Compound 103 appeared as an outlier in all the models, while compounds 104 and 105 appear as outliers in four models (in 1, 2, 3, and 5 and in 2, 3, 4, and 5, respectively). The fact that these compounds appear as outliers in all or most of the models raises two important issues. First, none of the descriptors considered for the study are able to capture the exact structural feature of these compounds that may be related to biological activity, and, second, the different modeling techniques adopted to derive the models are inadequate in relating these compounds to the observed property.

We consider the example of compound 128, in order to highlight the structural uniqueness of a compound that might have caused it to be an outlier. The QSAR Models 1−3 propose that the presence of any positive ionizable function and the absence of a hydrogen bond acceptor group at R_2 favors ligand binding. But compound 128 lacks a positive ionizable function and also possesses a hydrogen bond acceptor function; therefore, one would assume this compound to be less potent. In contrast, it turns out to be one of the most active compounds in the present study (pIC_{50} = 8.40). In Models 4 and 5, where these variables do not appear in combination, this molecule does not separate as an outlier. The point worth mentioning to conclude the argument upon outliers is, in the current study only few descriptors, which could be framed to a physicochemical interpretation, have been used. In light of this limitation and the restriction imposed on regression techniques to evolve only interpretable models might also be a cause for outliers. However, we believe that the chemical insights and the understanding regarding various aspects of protein−ligand interaction gained through the eyes of QSAR outweigh these shortcomings.

Molecular Docking and Prediction of Binding Mode. With the interest to determine the mode of inhibition of these ligands and to gain further confirmation of the physicochemical requirements as ascertained from QSAR, few key inhibitors that demonstrate the most characteristic protein−ligand interactions were docked into the ATP-pocket of CDK4 (Table 9). With visual inspection as our guide, we wished to select those poses that possessed at least two hydrogen bonding interactions with Val-96. All ATP targeted CDK2 inhibitors interact with Leu-83,[28] and, since the corresponding residue in CDK4 is Val-96, this hydrogen bonding appears to be critical.[29] In seven cases, of the nine ligands considered, the top ranked pose identified by GOLD conserved this H-bonding interaction between the ligand and Val-96, whereas in compounds 91 and 113, the fourth and sixth poses, respectively, were found to possess this key interaction and were selected for further investigation.

CDK4 Binding Site. The ATP pocket is substantially conserved across the three CDKs − 2, 4, and 6 (Table 10). Few differences exist between the CDK2 and CDK4 binding sites, replacement of Lys-89 with Thr-102 and substitution of Phe-82 and His-84 with His-95 and Asp-97. Also the hinge residue, Leu-83, in CDK2 is mutated to a very similar Val-96 in CDK4. In contrast the hinge residues of CDK4 are identical to that of CDK6. Both the cocrystallized structures, of ATP with CDK2 (1FIN) and ligand 113 in the ATP pocket of CDK6 (2EUF), display the typical kinase hydrogen bonding with hinge region. In the crystal structure of CDK2, ATP is hydrogen bonded to Leu-83, while ligand 113 is involved in a donor−acceptor dyad with Val-101 of CDK6. Also, the alignment of CDK4 to CDK6 and 2 indicated Val-96 to be the hinge residue capable of forming the hydrogen bond [Figure 2]. For these reasons, Val-96 was used to define the binding site in our docking experiments.

Comparison of CDK4 and CDK6. The binding site residues of CDK4 and CDK6 that form the major interacting sites for ligand 113[30] are listed in Table 10. Ligand 113 has also been cocrystallized with CDK6,[21] and as a consequence we have discussed in detail the critical residues in CDK 4 and 6 that form the core of the ATP pocket. On investigation, these residues were found to be almost identical (Table 10). Figure 8 illustrates the docked and the cocrystallized structure of 113 in CDK4 and CDK6, respectively. The figure also highlights the hydrogen bonding interactions between the ligand and protein. In both cases, the pyrido[2,3-d]pyrimidin-7-one core is held by two hydrogen bonds between the ligand and the hinge region of the protein. In CDK4, the amino (C2-NH) donates a hydrogen bond to the oxygen of Val-96, while the N3 nitrogen accepts a hydrogen bond from the amide of Val-96. In CDK6, the hydrogen bonding is with the corresponding Val101 of the hinge region. In our docked conformation we also observed a hydrogen bonding interaction between the pyridine nitrogen of the ligand and His-95 of CDK4, but no such interaction was observed in the cocrystallized structure of CDK6 with 113. The positively charged nitrogen of piperazine, which is exposed at the mouth of ATP-pocket, is stabilized by the hydrogen bonding

Table 9. Scores and the Interaction Energy Values Calculated for Compounds Docked into the Protein

	1	88	89	90	91	102	104	113	128
Chemscore	26.561	33.939	33.549	31.456	30.695	32.114	31.472	32.161	29.760
Ludi3	503	612	659	619	639	681	652	669	653
Ludi3_NHBOND	2	3	4	4	4	4	4	4	3
Ludi3_HBONDSCO	106	113	172	186	181	123	159	155	114
Ludi3_NLIP	56	54	55	52	53	59	55	58	59
Ludi3_LIPSCO	253	355	343	289	314	478	349	370	395
interaction energy (kcal/mol)									
E_{vdW}	−55.185	−68.745	−69.669	−63.538	−67.563	−80.044	−74.445	−77.258	−77.324
E_{Elect}	−11.561	−106.669	−108.863	−106.112	−108.401	−108.740	−105.605	−105.070	−9.871
E_{Total}	−66.746	−175.414	−178.532	−169.650	−175.964	−188.784	−180.050	−182.328	−87.195

Table 10. Comparison of Hinge Residues (CDK 2, 4, and 6) and Critical Residues that Form the Active Site in CDK4 and 6[a]

Residues that form the hinge region			Comparison of active site residues	
CDK4	CDK6	CDK2	CDK4	CDK6 (2EUF)
F93	F98	F80	I12	I19
E94	E99	E81	V20	V27
H95	H100	F82	A33	A41
V96	V101	L83	K35	K43
D97	D102	H84	G56	G61
Q98	Q103	Q85	V72	V77
D99	D104	D86	E144	Q149
T102	T107	K89	F159	F164

[a] Refer Figure 2 for color coding.

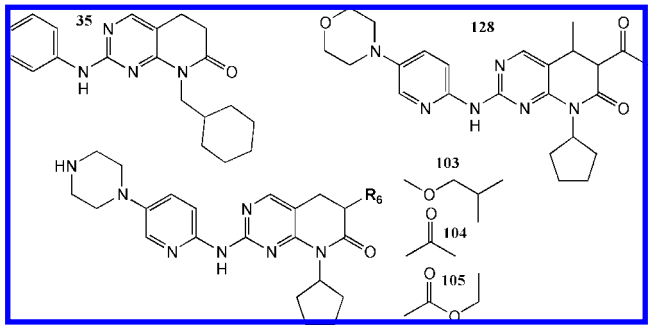


Figure 7. Structures of the few compounds that appeared as common outliers in the QSAR models.

interaction with Thr-102. The Asp-99 residue has been proposed to stabilize this positive charge on the ligand[29] and is specifically addressed in subsequent discussions. The one notable difference between the docked (in CDK4) and bound cocrystallized conformation (in CDK6) of ligand 113 is the orientation of the R_2 fragment at the entrance of the ATP pocket. In CDK4 the positive ionizable group orients toward Thr-102, while in CDK6 it deviates away from Thr-107 and bends toward Gln-149.

In CDK2, Phe-80 is often referred to as the "gate keeper" residue for it guards ATP from accessing the hydrophobic residue at the back of the ATP pocket. Interestingly it also offers an additional interaction point for the ligand as observed from the cocrystallized structure of CDK2 with roscovitine (PDB ID: 2A4L).[32,33] Extending the same scenario to CDK6 and CDK4, Phe-98, and Phe-93, respectively, should be the residues capable of providing this hydrophobic interaction point for ligands capable of extending a rightly oriented hydrophobic moiety. In 113 the R_5 methyl fragment and the R_6 acetyl moiety exploit this advantage. In CDK4, the methyl extends toward a small hydrophobic region formed by Val-20, Ala-33, Val-72, and Phe-93, very similar to what can be observed for CDK6 (shown in Table 10). The acetyl group at R_6 is stabilized by hydrophobic interactions with Phe-93, Val-72, and by the side chain of Lys-35. In CDK6 a hydrogen bonding interaction is observed between the carbonyl oxygen of acetyl and the backbone amide of Asp-163. In the docked conformation, the R_6 acetyl fragment orients toward the Asp-158 residue of CDK4, although a clear hydrogen bonding interaction was not observed. The R_8 cyclopentyl fragment is encapsulated from the solvent by hydrophobic residues Ile-12, Val-20, Leu-147, Ala-157, and the side chain of Glu-144 (Table 11).

Table 12 summarizes the key hydrogen bonding interaction observed between the docked ligands and the modeled CDK4. All ligands interact with the hinge region through a couple of hydrogen bonds with the backbone amide and carbonyl oxygen of Val-96. Compounds 88−91, 102, 104, and 113 that possess a positively charged nitrogen are able to make hydrogen bonds with Thr102. In compounds 89−91, 102, 104, 113, and 128 that possess a pyridine substituent at R_2, an additional hydrogen bonding is observed between the ϵ-NH of His-95 and pyridine nitrogen of the ligand [Figure 9].

(a) R_2. On the basis of QSAR equations (Models 1−3), the presence of indicator variable I_p with a positive coefficient implies that a positive ionizable group at R_2 makes a favorable contribution toward ligand binding. For all, but 1 and 128, the electrostatic interaction energy between the protein−ligand was quite high, and, specifically for ligand 88, the electrostatic interaction value between Asp-99 and ligand was determined as −64.865 kcal/mol. Hence the major intermolecular electrostatic protein−ligand interaction appears to have been yielded from the ionic contribution from the positively charged amino group of the ligand and the negatively charged Asp-99 of the protein. Docking studies by the Cyclacel group[29] validated the necessity to include a geometrically appropriate positively ionizable group charged at physiological pH at the R_2 region that would be involved in Coulombic electrostatic interaction with Asp-99. If this was to be strictly followed, one expects a compound with the absence of a positively ionizable function at the mouth of the ATP pocket to be less potent. In contrast to the proposal put forward, the absence of such a positive ionizable group does not seem to hinder biological activity. Even morpholine and piperidine, unlike piperazine, not charged at physiological pH, are potent inhibitors of CDK4, and this demands for further investigations to articulate this mecha-

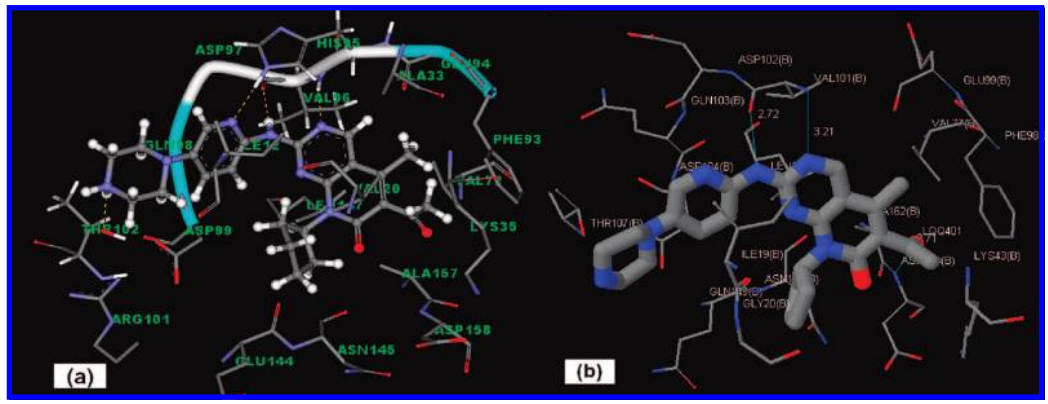


Figure 8. (a) The predicted binding mode of compound 113. The backbone of the hinge region is highlighted. (b) The cocrystallized protein−ligand conformation of compound 113 within ATP pocket of CDK6 (PDB ID: 2EUF) (acquired from PDBsum Web site[31]).

Table 11. Results Gathered from Docking Studies

	residues anchoring the fragments	nature of critical interactions	conclusions derived from docking
R_2	Ile12, Val96, His95, Asp99, and Thr102	hinge hydrogen bonding, Coulombic interaction with Asp-99	Pyridine shows H-bonding with His95.
R_5	Val-20, Ala-33, Val-72, and Phe-93	only hydrophobic	Only methyl seems acceptable.
R_6	Leu-60, Val-72, Phe-93, Ala-157 Phe-159 and Lys-35, Glu-56 and Asp-158	hydrophobic and polar	Bulky residues occupy a hydrophobic pocket at the back of ATP pocket. QSAR campaigns for polar substituents over hydrophobic.
R_8	Ile-12, Val-20, Leu-147, and Ala-157	hydrophobic	Only hydrophobic fragments tolerated.

Table 12. Hydrogen Bonding Interactions between the Ligand and Protein as Observed from Docking

mol. id	His95	Val96	Thr102	total
1	0	2	0	2
88	0	2	1	3
89	1	2	1	4
90	1	2	1	4
91	1	2	1	4
102	1	2	1	4
104	1	2	1	4
113	1	2	1	4
128	1	2	0	3

nism. We also noted that ligand 89 interacted with protein better than 88 (−178.532 versus −175.414). As depicted in Figure 9, compound 89 shows an additional hydrogen bonding interaction with the imidazole side chain of His-95. The stability of this hydrogen bond and its contribution toward stabilizing the protein−ligand complex is reflected in terms of the tremendous increase in Ludi3_HBONDSCO for 89 when compared to 88. Also in cases where a pyridine ring was present as a substituent at R_2, this hydrogen bonding interaction with His-95 was retained (Table 12). The result gathered corroborates the observation made by the original authors[10] regarding the selectivity shift on replacing benzene by pyridine. A more comprehensive study is underway to investigate this interaction and how much it contributes to selectivity.

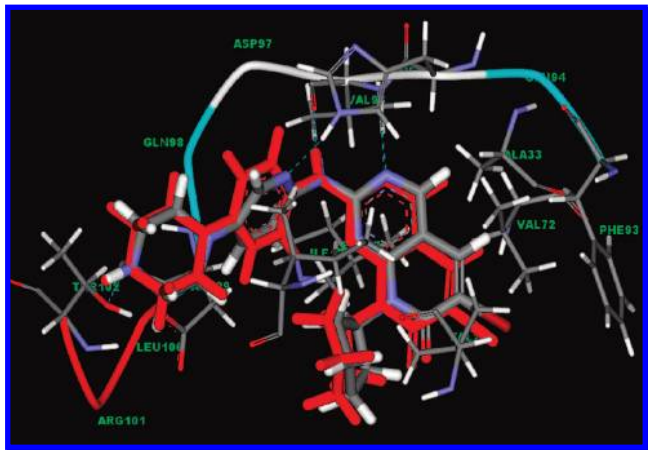


Figure 9. Compounds 89 (CPK color scheme) and 88 (red) shown within the ATP pocket of CDK4. Note the hydrogen bonding interaction of 89 with His95.

(b) R_5. The docking experiments convey that methyl as a substituent at R_5 is favorable. Comparing the docking results of compounds 113 and 104, with the former possessing a methyl at R_5, both were found to overlap nicely within the binding pocket with a methyl extending toward a small hydrophobic pocket of Val-20, Ala-33, Val-72, and Phe-93. Calculation of the intermolecular interaction energy showed compound 113 to interact better than compound 104 (−182.328 as against −180.050). The presence of a methyl at R_5 expectedly enhances the E_{vdW} contribution of the ligand (−77.258 versus −74.445). The Ludi3_LIPSCO for 113 and 104 (370 versus 349) also support this finding. Typically, the methyl contributes to the overall binding by enhancing the E_{vdW} input to the overall intermolecular (protein−ligand) interaction energy. The methyl therefore seems to be well tolerated within the ATP pocket. QSARs predict any substitution at this position to be detrimental for inhibition. At this region, docking experiments convey a methyl is tolerated well in the ATP pocket, while QSAR predicts even a methyl to be unfavorable as a substitution. Although the docking results advocate for a methyl at R_5, they do not in any way contradict the conclusion drawn from QSAR. From the docking results it can be interpreted that only methyl as a substituent fits the small hydrophobic cavity at this region.

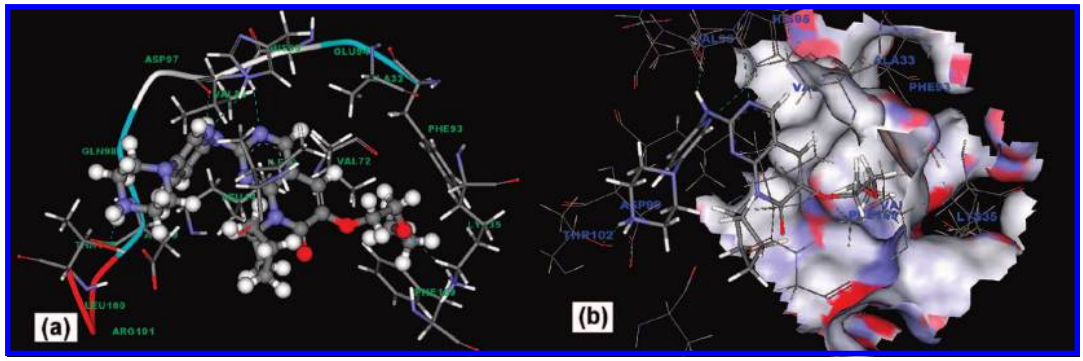


Figure 10. (a) Compound 102 in its docked conformation. (b) The protein residue that encapsulates the long R_6 chain of 102, represented as a solvent accessible surface. The R_6 substituent is shown here penetrating into the cavity at the back of ATP pocket.

(c) R_6. This region seems to be very intriguing considering (i) substitution at this region seems to be essential (model 2), (ii) preferable substitution being a low logP fragment ($\leq$0.210, Models 4 and 5), and (iii) tolerance toward bulky substituents (Model 3). These results taken together suggest the existence of a cavity that is able to accommodate these bulky substituents. In QSAR equations, indicator variable I_6 (Model 2) and Verloop parameter VB1 (R_6) (Model 3) augment upon the necessity to include substitution at position R_6 as well as their ability to accommodate bulky substitution. The QSAR equation also indicates that hydrophobic fragments at this position are detrimental toward activity (Models 1, 2, and 3). In order to determine the position of such bulky fragments within the ATP-pocket we docked compound 102 into the ATP pocket of CDK4 [Figure 10(a)]. Compound 102 possessed one of the biggest substitutions studied at this region hence preferred for validation by docking. We observed that the bulky fragment enters into the hydrophobic groove formed by Leu-60, Val-72, Phe-93, Ala-157, and Phe-159 [Figure 10(b)]. Three polar amino acids, Lys-35, Glu-56, and Asp-158, stand at the entrance of this cavity. Hence an optimal fragment at this position should also possess some polar functions that favorably interact with any of these side chains. In compound 102 the ethoxy fragment performs this role perfectly. The ethoxy oxygen is within hydrogen bonding distance with Lys35, while the $-CH_2CH_3$ occupies the hydrophobic pocket.

In the present class of compounds, a variety of substituents at R_6 (compounds 89 and 93−103) seem to be well accommodated in the binding pocket without any substantial loss in activity (except in the case of 103). This observation made us question the occupancy of these bulky substituents within the protein. Analyses of docking results on compounds 102 and 104 resulted in unleashing this ambiguity. Compound 102 with its long substituent at R_6 enters a narrow channel at the back of ATP, while the small acetyl just hangs its methyl at this entrance. Since this channel is narrow, compounds with branched side chains, like 103 (R_6: $-OCH_2{}^iPr$), are unable to penetrate this cavity. The hydrophobic interaction prevailing at this region and its contribution to binding is reflected in the intermolecular interaction energy and the Ludi score. The tremendous increase in Ludi3_LIPOSCO for compound 102 as compared to 104 (478 against 349) authenticates the hydrophobic nature of this pocket. Also the differences in nonbonded van der Waals interaction energy (E_{vdW}: −80.044 against −74.445) and the Ludi3 score correspond well with the observed potency differences for compounds 102 and 104 (102 is more potent than 104). Our QSAR study strongly advocates for polar substituents over hydrophobic. The docking experiments, although they do not contradict this proposal, seem to give a better assessment of protein−ligand interaction at this region than QSAR. The hidden hydrophobic pocket explored by bulky long chain substituents is better understandable through docking experiments. This in principle could be one of the reasons for the outliers 103, 104, and 105. It seems that the statistical tools are unable to extract this information from the descriptors included in the study. Also the docking experiments convey a dual nature of substituents to be more favorable. In conclusion, the docking experiments confirm the tolerance of substituents containing both hydrophobic and polar fragments, which QSAR fails to capture.

(d) R_8. In order to highlight the importance of a critically bulky hydrophobic moiety at this region, comparison is drawn between compounds 90 and 91; the former holds a cyclopropyl, whereas the latter possess a cyclopentyl at R_8. Compound 91 enjoys a better hydrophobic interaction with the receptor as compared to 90 (E_{vdW}: −67.563 versus −63.538, Ludi3_LIPSCO: 314 against 289). The adequate bulkiness, vital for these hydrophobic fragments, can be argued to be essential to interact with amino acids at this region. Also on comparison of compounds 1, 23, and 24, the methyl in 23 is too small to enjoy the full lipophilic interactions offered by the amino acids at this region, whereas consequent substitution to more bulky ethyl and isopropyl (compounds 1 and 24, respectively) seems to boost these interactions leading to more energetically favorable ligand binding further advocating the proposal put forward.

CONCLUSIONS

QSARs have been derived that are able to quantitatively correlate physicochemical parameters of molecules to CDK4 inhibitory activity. The structural characteristics derived from these models were validated and interpreted with the aid of a homology modeled 3D structure of CDK4. The QSARs not only were statistically significant but also revealed the following requirements for CDK4 inhibition based on the pyrido[2,3-d]pyrimidin-7-one core: (i) R_2 - bulky hydrophilic fragments with a positive ionizable group, (ii) R_5 - bulky substitution disfavored, small hydrophobic group, methyl is tolerated, (iii) R_6 - substitution at this position is favorable, and the preferred fragment at this region would be a substituent that is polar in nature, and (iv) R_8 - only hydrophobic fragments preferred. Docking performed on a homology model of CDK4 provided a mechanistic interpre-

tation of the derived QSAR as well as the molecular basis governing these protein−ligand interactions. Docking identified a unique hydrogen bonding between the side chain of His-95 and the pyridine nitrogen at R_2. It was also possible to understand the importance of methyl at R_5 with the aid of docking. The docking study also divulges that bulky substituents at R_6 occupy a hydrophobic groove at the back of the ATP pocket. The hydrophobic fragments at R_8 need to be critically bulky to savor good interaction with the protein. We strongly believe that the insights provided in this study will further help in understanding the specific nature of protein−ligand interactions that are crucial toward inhibition of CDK4, which, in turn, should help to develop potential inhibitors for this target.

ACKNOWLEDGMENT

N.M.M. thanks CSIR for the award of fellowship (SRF), and N.G. is grateful for the financial support from Mission Mode Program of CSIR (CMM 0017).

Supporting Information Available: Tables of structures selected for the study with their pIC_{50} values (Table 1), predicted biological activity (pIC_{50}) with the five models (Table 5), and correlation plot of descriptors that appear in any of the five models (Table 7) and Model 6. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Morgan, D. O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu. Rev. Cell Dev. Biol.* **1997**, *13*, 261–291.
(2) Malumbres, M.; Barbacid, M. Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.* **2005**, *30*, 630–641.
(3) Kasten, M. M.; Giordano, A. pRb and the Cdks in apoptosis and cell cycle. *Cell Death Differ.* **1998**, *5*, 132–140.
(4) Ortega, S.; Malumbres, M.; Barbacid, M. Cyclin D-dependent kinases INK4 inhibitors and cancer. *Biochim. Biophys. Acta* **2002**, *1602*, 73–87.
(5) Ortega, S.; Malumbres, M.; Barbacid, M. Cell cycle and cancer: the G1 restriction point and the G1/S transition. *Curr. Genomic.* **2002**, *3*, 245–263.
(6) Landis, M. W.; Pawlyk, B. S.; Li, T.; Sicinski, P.; Hinds, P. W. Cyclin D1-dependent kinase activity in murine development and mammary tumorigenesis. *Cancer Cell* **2006**, *9*, 13–22.
(7) Yu, Q.; Sicinska, E.; Geng, Y.; Ahnstrom, M.; Zagozdzon, A.; Kong, Y.; Gardner, H.; Kiyokawa, H.; Harris, L. N.; Stal, O.; and Sicinski, P. Requirement for CDK4 kinase function in breast cancer. *Cancer Cell* **2006**, *9*, 23–32.
(8) Barvian, M.; Boschelli, H. D.; Cossrow, J.; Dobrusin, E.; Fattaey; A.; Fritsch, A.; Fry, D.; Harvey, P.; Keller, P.; Garret, M.; La, F.; Leopoid, W.; McNamara, D.; Quin, M.; Kallmeyer, S. T.; Toogood, P. L.; Wu, Z.; Zhang, E. Pyrido[2,3-d]pyrimidin-7-one inhibitors of cyclin-dependent kinases. *J. Med. Chem.* **2000**, *43*, 4606–4616.
(9) VanderWel, S. N.; Harvey, P. J.; McNamara, D. J.; Repine, J. T.; Keller, P. R.; Quin, J.; Booth, J. R.; Elliott., W. L.; Dobrusin, E. M.; Fry, D. W.; Toogood, P. L. Pyrido[2,3-d]pyrimidin-7-ones as specific inhibitors of cyclin-dependent kinase 4. *J. Med. Chem.* **2005**, *48*, 2371–2387.
(10) Toogood, P. L.; Harvey, P. J.; Harvey, P. J.; Repine, J. T.; Sheehan, D. J.; VanderWel, S. N.; Zhou, H.; Keller, P. R.; McNamara, D. J.; Sherry, D.; Zhu, T.; Brodfuehrer, J.; Choi, C.; Barvian, M. R.; Fry, D. W. Discovery of a potent and selective inhibitor of cyclin-dependent kinase 4/6. *J. Med. Chem.* **2005**, *48*, 2388–2406.
(11) A product of Accelrys Inc. http://www.accelrys.com.
(12) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000–1008.
(13) A product of Chemical Computing Group. http://www.chemcomp.com.
(14) Kadam, R. U.; Roy, N. Cluster analysis and two-dimensional quantitative structure-activity relationship (2D-QSAR) of Pseudomonas aeruginosa deacetylase LpxC inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5136–5143.
(15) Stanton, D. T. On the Physical Interpretation of QSAR Models. *J. Chem. Inf. Model.* **2000**, *43*, 1423–1433.
(16) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
(17) Verloop, A.; Hoogenstraaten, W.; Tipker, J. Development and Application of New Steric Substituent Parameters in Drug Design. In *Drug Design*; Ariens, E. J., Eds.; Academic Press: New York, 1976; Vol. 7, pp 165−207.
(18) pKa calculations were performed using MarvinBeans 4.1.2, 2006, ChemAxon. http://www.chemaxon.com.
(19) Rogers, D.; Hopfinger, A. Application of genetic function approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
(20) A product of SPSS Inc. http://www.spss.com.
(21) Lu, H.; Schulze-Gahmen, U. Toward understanding the structural basis of cyclin-dependent kinase 6 specific inhibition. *J. Med. Chem.* **2006**, *49*, 3826–3831.
(22) Jeffrey, P. D.; Russo, A. A.; Polyak, K.; Gibbs, E.; Hurwitz, J.; Massagué, J.; Pavletich, N. P. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **1995**, *376*, 313–320.
(23) Laskowski, R. A.; MacArthur, M. W.; Moss, D.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
(24) Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
(25) Böhm, H. J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from the de novo design or 3D database search programs. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309–323.
(26) Verma, R. P.; Hanch, C. An approach toward the problem of outliers in QSAR Bioorg. *Med. Chem.* **2005**, *13*, 4597–4621.
(27) Kim, K. H. Outliers in SAR and QSAR: Is unusual binding mode a possible source of outliers. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 63–86.
(28) Mascarenhas, N. M.; Ghoshal, N. *Eur. J. Med. Chem.* doi:10.1016/j.ejmech.2007.10.016, 2007.
(29) McInnes, C.; Wang, S.; Anderson, S.; O'Boyle, J.; Jackson, W.; Kontopidis, G.; Meades, C.; Mezna, M.; Thosma, M.; Wood, G.; Lane, D. P.; Fischer, P. M. Structural determinants of CDK4 inhibition and design of selective ATP competitive inhibitors. *Chem. Biol.* **2004**, *11*, 525–534.
(30) Fry, D. W.; Harvey, P. J.; Keller, R. P.; Elliot, W. L.; Meade, M.; Trachet, E.; Albassam, M.; Zheng, X.; Leopold, W. R.; Pryer, N. K.; Toogood, P. L. Specific inhibition of cyclin-dependent kinase 4/6 by PD 0333191 and associated antitumor activity in human tumor xenografts. *Mol. Can. Ther.* **2004**, *3*, 1427–1438.
(31) Laskowski, R. A.; Hutchinson, E. G.; Michie, A. D.; Wallace, A. C.; Jones, M. L.; Thornton, J. M. PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends. Biochem. Sci.* **1997**, *22*, 488–490.
(32) De Azevedo, W. F.; Leclerc, S.; Meijer, L.; Havlicek, L.; Strnad, M.; Kim, S. H. Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human cdk2 complexed with roscovitine. *Eur. J. Biochem.* **1997**, *243*, 518–526.
(33) Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. Protein Kinase Inhibitors: Insights into Drug Design from Structure. *Science* **2004**, *303*, 1800–1805.

CI8000343