ISSN 0975-6299 Vol.1/Issue-4/Oct-Dec.2010

International Journal of Pharma and Bio Sciences

PHARMACOPHORE MAPPING – AN APPROACH FOR DESIGNING CYCLIN DEPENDENT KINASE 2 INHIBITORS FOR CANCER TREATMENT

[1]ASHISH SETHI AND [2*]SACHDEV YADAV

[1] M.S Pharm II Sem (Pharmacoinformatics), NIPER Hajipur, India.
[2] Assistant Professor (Pharmacology), Department of Pharmacy Banasthali University, P.O. Banasthali Vidyapith 304022, Rajasthan, India.

**Corresponding Author* sachdev_y@yahoo.com

ABSTRACT

CDKs are serine/threonine kinases that play pivotal roles in cell cycle progression. Their timely activation and deactivation drives the cell through different stages of the cell cycle. The regulation of CDK activity is achieved by their association with cyclins, specific CDK inhibitors, their state of phosphorylation and ubiquitin mediated degradation. Several CDKs and cyclins have been identified, though only CDKs 1, 2, 4 and 6 are known to intervene in the cell cycle machinery.Majority of inhibitors target the ATP pocket of CDK2 and are labeled ATP-competitive inhibitors. Few CDK2 inhibitors have entered clinical trials and are displaying encouraging results.Virtual Screening (VS) is the computational analogue of biological screening. The main objective of VS is to help medicinal chemist filter out in actives from a library of compounds before going ahead for synthesis. An alternate approach to the above scheme involves extraction of common 3D features from a set of ligands, known as pharmacophore.

KEY WORDS

CDKs, pharmacophore, anti- cancer drugs.

1. INTRODUCTION

Cell cycle is highly controlled and regulated by various mechanisms in mammalian cell The outcome of a successful cell cycle is a healthy replication of DNA into two daughter cells. The cell cycle is divided into four phases, G1, S, G2 and M phase. Between these phases are the checkpoints, which control the commitment of cell to further progress. The two extensively studied checkpoints appear at the G1/S and G2/M boundary. When checkpoint arrest controls are compromised, it leads to initiation of S phase or mitosis despite cellular damage leading to cancerous cells. Hence, molecules that act as decision makers at these checkpoints are good targets for the treatment of cancer. For this reason, Cyclin-Dependent Kinases (CDKs) are prospective targets [1].

CDKs are serine/threonine kinases that play pivotal roles in cell cycle progression [2]. Their timely activation and deactivation drives the cell through different stages of the cell cycle. The regulation of CDK activity is achieved by their association with cyclins, specific CDK inhibitors, their state of phosphorylation and ubiquitin mediated degradation. Several CDKs and cyclins have been identified, though only

CDKs 1, 2, 4 and 6 are known to intervene in the cell cycle machinery [3, 4]. CDK2 is involved in cell cycle progression at two different stages; CDK2, activated by Cyclin E during G1 phase is the checkpoint for G1/S phase transition, while binding to Cyclin A during the G1/S, drives the cell through the S phase. The substrates of CDK2/Cyclin A (CDK2A) are being investigated and previous studies reveal that CDK2A phosphorylates tumor suppressor pRb and related proteins p107 and p130, transcription factors and CDC6 [3]. The protein inhibitors of these CDKs are down regulated in most of the cancer cells. For this reason considerable amount of interest has evolved to develop inhibitors that target CDK2 for treating cancer. Majority of inhibitors target the ATP pocket of CDK2 and are labeled ATP-competitive inhibitors. Few CDK2 inhibitors have entered clinical trials and are displaying encouraging results [5].

Virtual Screening (VS) is the computational analogue of biological screening [4]. The main objective of VS is to help medicinal chemist filter out in actives from a library of compounds before going ahead for synthesis. An alternate approach to the above scheme involves extraction of common 3D features from a set of ligands, known as pharmacophore,followed by 3D database search to retrieve hits that satisfy these features.

The goal of the present study was to explore the possibility of adopting feature-shape query built from the biologically active conformation of the ligand for VS. protein ligand interactions responsible for inhibiting CDK2A has been gained that may be used as a guide in designing potent anti- cancer drugs.

2. Cyclin dependent kinase

2.1 Type

Table 1
Showing different kind of CDKs with their regulatory proteins [2]

S.No.	CDK	Regulatory proteins
1.	CDK1	yclin A, Cyclin B
2.	CDK2	Cyclin A, Cyclin E
3.	CDK3	
4.	CDK4	Cyclin D1, Cyclin D2, CyclinD3
5.	CDK5	Cyclin 5R1, Cyclin 5R2, CyclinL5
6.	CDK6	Cyclin D1, Cyclin D2, CyclinD3
7.	CDK7	Cyclin H
8.	CDK8	Cyclin C
9.	CDK9	Cyclin T1, CyclinT2a
		CyclinT2b, Cyclin K
10.	CDK10	
11.	CDK11/CDC2L2	Cyclin L

2.2 Regulation

A cyclin-CDK complex can be regulated by several kinases and phosphatases, including Wee, and CDK-activating kinase (CAK), and Cdc25. CAK adds an activating phosphate to the complex, while Wee adds an inhibitory phosphate, the presence of both activating and inhibitory phosphate added by Wee, rendering the complex active. CDK feeds back on Wee and Cdc25 to inhibit and enhance their respective activities.(as shown in **Figure 1)**

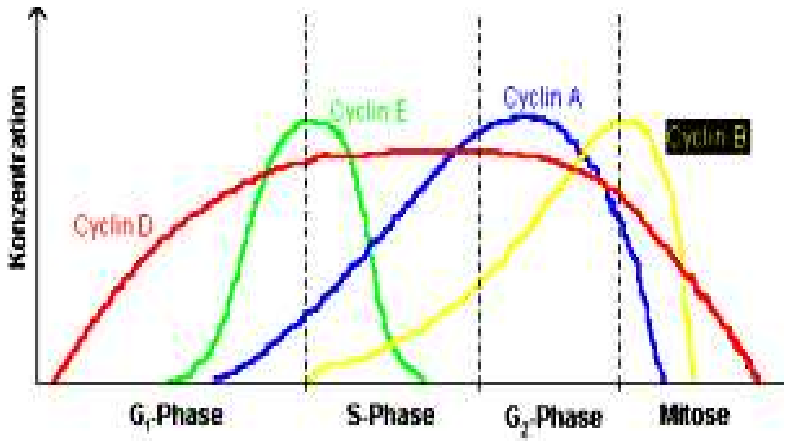


Figure 1
Level of various cyclin during various phases of cell cycle [1]

Cyclin D is the first cyclin produced in the cell cycle, in response to extracellular signal (e.g. growth factor). Cyclin D binds to existing CDK4, forming the active cyclin D-CDK4 complex. Cyclin D-CDK4 complex in turn phosphorylates the retinoblastoma susceptibility protein (Rb). The hyperphosphorylated Rb dissociate from the E2F/DP1/Rb complex (which was bound to the E2F responsive gene, effectively "blocking" them from transcription), activating E2F. Activation of E2F results in transcription of various genes like cyclin E, cyclin A, DNA polymerase, thymidine kinase, etc. Cyclin E thus produced binds to CDK2, forming the cyclin E-CDK2 complex, which pushes the cell from G_1 to S phase (G_1 /S transition). Cyclin B along with CDC2 forms the cyclin B-CDC2 complex, which initiates the G_2/M transition [8]. Cyclin B-CDC2 complex activation causes breakdown of nuclear envelop and initiation of prophase, and subsequently, its deactivation causes the cell to exit mitosis.

2.3 Role of CDKs in cell cycle

Upon receiving a pro-mitotic extracellular signal, G_1 cyclin-CDK complexes become active to prepare the cell for S phase, promoting the expression of transcription factors that in turn promote the expression of S cyclins and of enzymes required for DNA replication. The G_1 cyclin-CDK complexes also promote the degradation of molecules that function as S phase inhibitors by targeting them for ubiquitination. Once a protein has been ubiquitinated, it is targeted for proteolytic degradation by the proteasome. Active S cyclin-CDK complexes phosphorylate proteins that make up the pre-replication complexes assembled during G_1 phase on DNA replication origins. The phosphorylation serves two purposes: to activate each already-assembled pre-replication complex, and to prevent new complexes from forming. This ensures that every portion of the cell's genome will be replicated once and only once. The reason for prevention of gaps

in replication is fairly clear, because daughter cells that are missing all or part of crucial genes will die. However, for reasons related to gene copy number effects, possession of extra copies of certain genes would also prove deleterious to the daughter cells. Mitotic cyclin-CDK complexes, which are synthesized but inactivated during S and G_2 phases, promote the initiation of mitosis by stimulating downstream proteins involved in chromosome condensation and mitotic spindle assembly. A critical complex activated during this process is an ubiquitin ligase known as the anaphase-promoting complex (APC), which promotes degradation of structural proteins associated with the chromosomal kinetochore. APC also targets the mitotic cyclins for degradation, ensuring that telophase and cytokinesis can proceed. Interphase generally lasts at least 12 to 24 hours in mammalian tissue. During this period, the cell is constantly synthesizing RNA, producing protein and growing in size.(refer **Figure 2**)

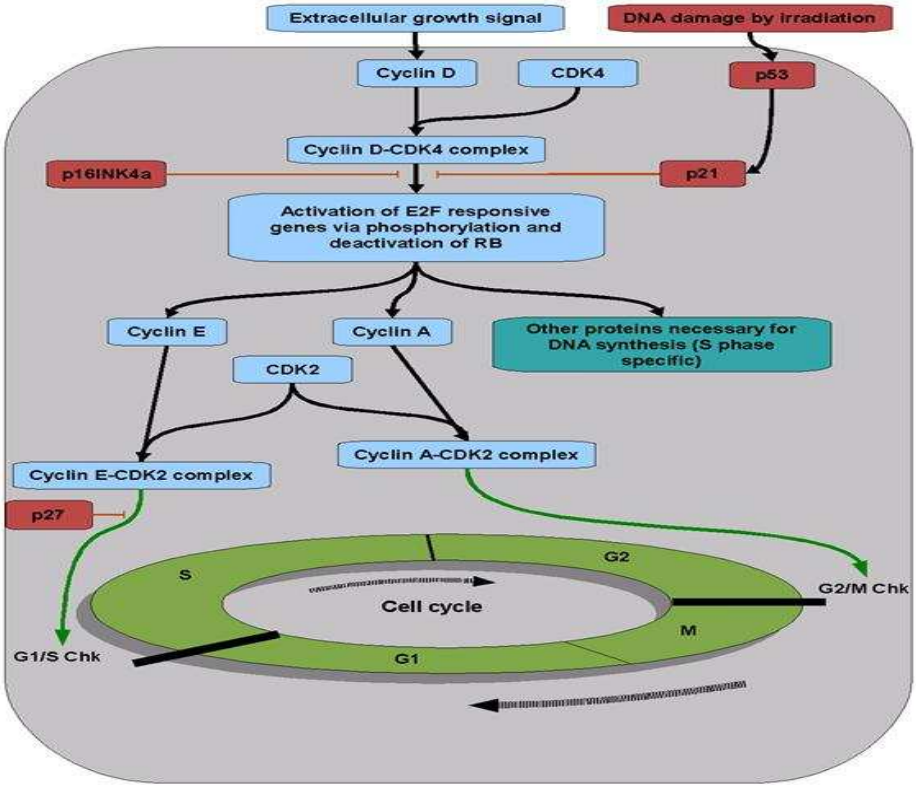


Figure 2
Role of CDKs and cyclins in regulation of cell cycle

2.4 CDK inhibitors mammals

Mammals possess two classes of CKIs (**Table 2**) which differ in structure, mechanism of inhibition, and specificity. One class, the p21 family, comprises $p21^{CIPI/WAF1}$, the product of a p53-regulated gene and the first mammalian CKI to be discovered, and two other CKIs, $p27^{KIP1}$ and $p5^{KIP2}$ [7]. These inhibitors have a conserved amino-terminal 60-residue domain responsible for kinase binding and inhibition and preferentially inhibit CDKs of the GI/S phase.

Table 2
Classification of CDK inhibitor [7]

Inhibitor	Primary target	Chromosomal location	Regulators	Comments
		Ankyrin family		
$P15^{INK4b}$	CDK 4-6	9p21	TGFβ	Frequently deleted in glioblastoma[26]
$P16^{INK4a}$	CDK 4-6	9p21	?	Tumor suppressor (melanoma) [30,31]
$P17^{INK4c}$	CDK 4-6	1p32	?	Not yet associated with cancer
$P18^{INK4d}$	CDK 4-6	19p13	?	Not yet associated with cancer
		Dual specificity family		
$P21^{CIP1/W\ AF1}$	CDK 2, 3, 4, 6	6p21	P53, TGFβ	Functions in G1 checkpoint
			MyoD	Possible tumor suppressor in prostrate
$P27^{KIP1}$	CDK 2, 4, 6	12p12-13	Rapamycin	Not yet associated with cancer
$P57^{KIP2}$	CDK2, 3, 4	11p15.5	?	Candidate tumor suppressor for both sporadic and familial

The second class of CKIs, referred to as the INK4 (Inhibitor of CDK4) family, is comprised of ankyrin repeat proteins and includes p15, p16, p18 and p19 (**Table 2**) [6]. p16 and, to a lesser extent, p15 [26] are found mutated or deleted in certain types of human cancers and several of these p16 mutations have been shown to be non-functional in cell-cycle arrest and CDK4 inhibition. In contrast to the p21 family, INK4 family members are selective for complexes of cyclin D with CDK4 or CDK6, and are not found associated with active kinase complexes. INK4 homologs can associate tightly with either the monomer CDK subunit or the cyclin-bound form [7], unlike the dual-specificity class, for which association with CDKs is largely cyclin-dependent.

3. CDK and Cancer

Proper development of a multi cellular organism requires precise spatial and temporal control of cell proliferation. A large network of regulatory genes has evolved to specify when and where in the organism cells divide. This network is superimposed onto the basic cell-cycle regulatory machinery, at the center of which lies the cyclin-dependent kinases (CDKs). These kinases are positively regulated by cyclin association and are required for cell-cycle progression [6]. While much is known about how cells enter the cell cycle, relatively little is known about the strategies involved in exit from the cell cycle and maintenance of the non-proliferative state. This state is of critical importance because the vast majority of animal cells exist in a non-proliferative state throughout adult life. The inability to halt growth appropriately can lead to malformation during organism development, and to cancer. Thus, equally important to the correct execution of developmental programs is the arrest of growth once the program is complete. Additional control circuits, called checkpoints, operate during cell proliferation to maintain the order and timing of cell-cycle events. Checkpoints also monitor the integrity of DNA and mediate cell-cycle arrest and repair processes in response to DNA damage [7]. Given the central role of CDKs in cell-cycle progression, it was anticipated that mechanisms responsible for cell- cycle arrest would include alterations in CDK activity. Conversely, mechanisms leading to increased CDK activity would be expected to promote uncontrolled cell proliferation. The discovery of multiple classes of CDK inhibitors (CKIs) has provided new paradigms for understanding how cell proliferation is modulated in response to a variety of extracellular and intracellular signals. Moreover, the finding that some CKIs act as tumor suppressors has provided a direct link between tumor genesis and the loss of negative control pathways directly affecting CDK activity. The identification of CKIs in single-cell eukaryotes [4,5] in addition to mammals points to the universality of this mechanism for cell-cycle arrest. The field of CKI activity and cell proliferation is moving rapidly; more than 130 related papers have emerged since the discovery of CKIs in late 1993, accompanied by reviews summarizing various aspects of the field [8].

3.1 Importance of CDK2 inhibition

CDK2 is involved in cell cycle progression at two different stages; CDK2, activated by Cyclin E during G1 phase is the checkpoint for G1/S phase transition, while binding to Cyclin A during the G1/S, drives the cell through the S phase [9]. The substrates of CDK2/Cyclin A (CDK2A) are being investigated and previous studies reveal that CDK2A phosphorylates tumor suppressor pRb and related proteins p107 and p130, transcription factors and CDC6. The protein inhibitors of these CDKs are down regulated in most of the cancer cells. For this reason considerable amount of interest has evolved to develop inhibitors that target CDK2 for treating cancer. Majority of inhibitors target the ATP pocket of CDK2 and are labeled ATP-competitive inhibitors. Few CDK2 inhibitors have entered clinical trials and are displaying encouraging results [10].

3.2 Review of CDK2 involvement in cell cycle

Table 3

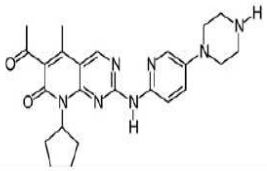
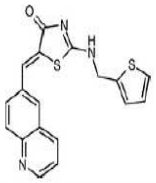
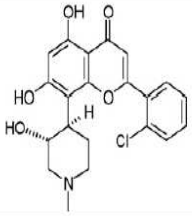
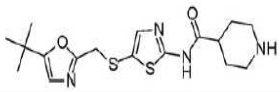
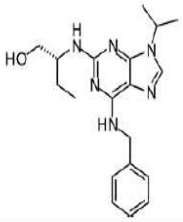
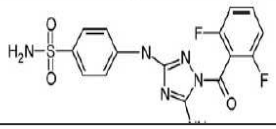
CDK involved in different phase of cell cycle with the cyclin protein binding to them [1].

Phase	Cyclin	CDK
G1	D, E	CDK4, CDK2,
S	A, E	CDK2
G2	A	CDK2
M	B	CDK1

4. Status of in vivo anti tumor activity

Table 4

CDK inhibitors in preclinical phase showing their specificity and phase of cell cycle [7]

CDK inhibitor	Chemical structure	CDK specificity	Phenotype	Clinical stage
PD0332991 (Pfizer)		CDK4, 6	G1 Arrest	Phase I (Lymphomas; Multiple Myeloma with Bortemizi
RO3306 (Hoffman-LaRoche)		CDK1	G2/M	Preclinical
Flavopiridol (Aventis)		CDK9	G1/S; G2/M; transcription inhibition	Numerous phase I and II; Phase I (CLL)
SNS-032 (Sunesis)		CDK2, 7,9	Cell cycle/ transcription inhibition	Phase I (advanced solid tumors; B-lympoid malignancies)
R-Roscovitine (Cyclacel)		CDK2, 7,9	G2/M, Transcription Inhibition	Phase I (various tumors; response in metastatic ovarian ca Phase II, lung and breast
JNJ-7706621 (Johnson and Johnson)		CDK1,2 (Aurora A and B)	G1 Delay, G2/M arrest, Endoreduplication	Preclinical

5. Computational study of CDK2 inhibitor

Pharmacophore modeling, including ligand-based and structure-based approaches, has become an important tool in drug discovery. However, the ligand-based method often strongly depends on the training set selection, and the structure-based pharmacophore model is usually created based on apo structures or a single protein–ligand complex, which might miss some important information. In this study, multicomplex-based method has been suggested to generate a comprehensive pharmacophore map of cyclin-dependent kinase 2 (CDK2) based on a collection of 124 crystal structures of human CDK2–inhibitor complex.

Our multi complex-based comprehensive pharmacophore map contains almost all the chemical features important for CDK2–inhibitor interactions[11]. A comparison with previously reported ligand-based pharmacophore has revealed that the ligand-based models are just a subset of our comprehensive map. Furthermore, one most-frequent-feature pharmacophore model consisting of the most frequent pharmacophore features was constructed based on the statistical frequency information provided by the comprehensive map. Validations to the most-frequent-feature model show that it can not only successfully discriminate between known CDK2 inhibitors and the molecules of focused inactive dataset, but also is capable of correctly predicting the activities of a wide variety of CDK2 inhibitors in an external active dataset. Obviously, this investigation provides some new ideas about how to develop a multicomplex-based pharmacophore model that can be used in virtual screening to discover novel potential lead compounds.

5.1 Pharmacophore mapping

Pharmacophore modeling method, as a key tool of computer aided drug design, has been widely used in the lead discovery and optimization [12].

The pharmacophore modeling method involves two aspects

Pharmacophore hypothesis generation 3D structural database search The latter is very mature right now and many advanced 3D database searching algorithms have been implemented in commercial programs. However, the pharmacophore hypothesis generation method is still under development although there are several commercial programs available presently[4]

The generation method of pharmacophore model can be classified into two categories:

Direct method - uses apo structure or receptor–ligand complex information (usually called receptor based or structure-based method).

Indirect method - uses a set of ligands that have been experimentally observed to interact with a specific biological target (called ligand-based method).

The ligand-based method has been very popular since for a long time just a limited number of protein structures or protein– ligand complex structures are available. It has been shown that the quality of the pharmacophore model generated by ligand based method can be affected significantly by two factors the conformation generation method and the selection of the training set molecules For the conformation generation method, systematic researches and large-scale survey works have been carried out by Langer, Gillette and others. It has been demonstrated that most of the conformation model generators are competent for the pharmacophore construction [12]. In order to establish a good pharmacophore model, specific guidelines are required to choose appropriate training set molecules, such as the activity range and the structural diversity of the selected compounds. The difference in the training set selection has a large influence on the final pharmacophore model. A possible case is that completely different pharmacophore models of ligands interacting with the same protein target could be generated with the use of the same

algorithm and program but different training set.

5.2 Pharmacophore modeling of CDK2 inhibitors

Hecker et al., Toba et al. and Vadivelan et al. have independently reported three ligand-based pharmacophore models for CDK2 inhibitors[12]. However, the three pharmacophore models are found to be different from each other in terms of the feature categories as well as the location constraint of features **Fig. 3**.

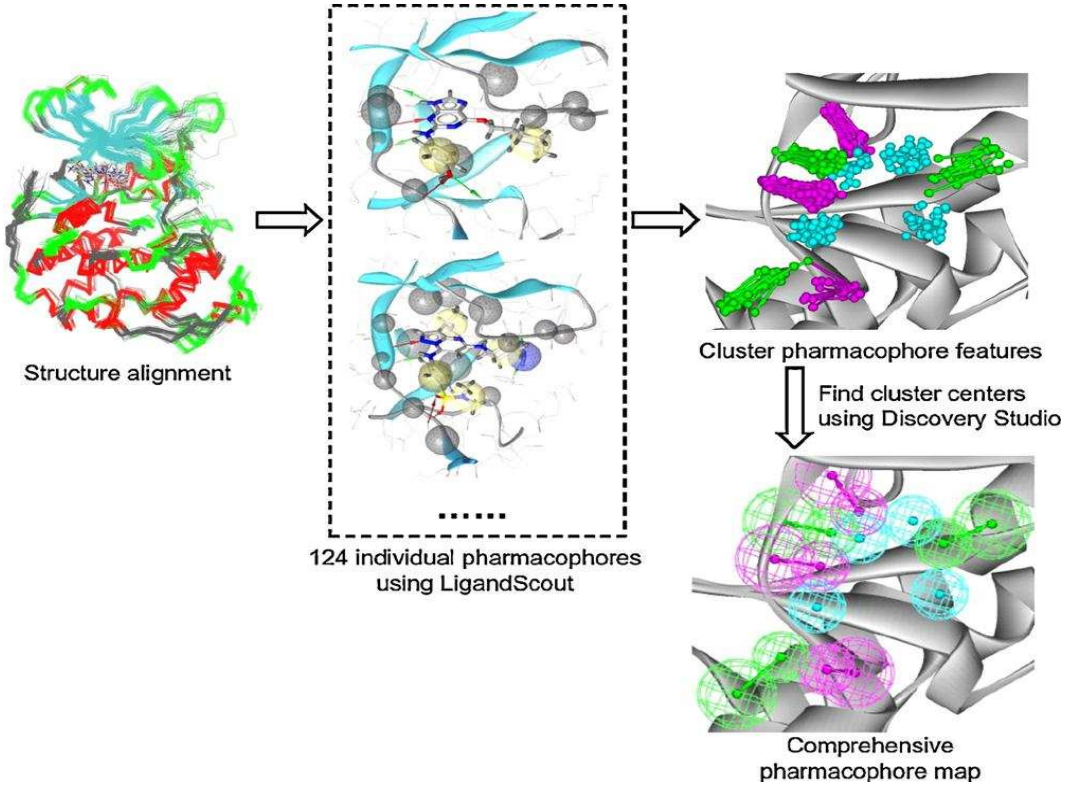


Figure 3
The ligand-based pharmacophore models of CDK2 inhibitors reported previously by (a) Hecker, (b) Toba and (c) Vadivelan. (A, hydrogen bond acceptor; B, hydrogen bond donor; H, hydrophobic feature; RA, ring aromatic feature.) [12].

These are mainly due to that different training set molecules were used in the three studies. A possible way to overcome this type of shortcomings is to adopt the structure-based pharmacophore modeling method, which just use the information of receptor or receptor–ligand complex [13].Usually structure-based pharmacophore models are generated based on apo structures or a single protein–ligand complex. However, in the case of the identical binding modes of different ligands, multiple complexes based pharmacophore would be the best since it allows for detecting all the protein–ligand interaction patterns and for evaluating the importance of each protein–ligand interaction. The resulting comprehensive pharmacophore map should contain much more information and be more accurate over others.

5.3 Generation of most-frequent-feature pharmacophore model

A comprehensive pharmacophore map was developed utilizing all available CDK2–inhibitor complex structures.

In this study a total of 124 crystal structures of CDK2–inhibitor complexes were taken from the protein data bank (PDB).

Then these structures were used to construct comprehensive pharmacophore map, which was used to compare previously reported ligand-based pharmacophore models, Finally, one most-frequent-feature pharmacophore model was generated based on the most frequent features of the comprehensive pharmacophore map.

Protein and ligand preparation

The structural data was used to generate a structure-based pharmacophore for the ATP-competitive inhibitors of CDK [12].

124 X-ray crystallography structures of CDK2 in complex with chemical inhibitors were obtained from the protein data bank (PDB).

Water molecules were removed due to their high mobility and to prevent complication.

Three multiple protein structure alignment tools, including Align3D in Modeller, MUSTANG, and 3DMA in Accelrys Discovery Studio, were used in this study to examine whether different algorithms influence the structural alignment or not.

The crystal structure with PDB code „„2C5O"", the one with the high crystallographic resolution and no missing residues, was arbitrarily taken as the reference structure

Generation of multicomplex-based comprehensive pharmacophore map

The whole process of generating the multicomplex-based comprehensive pharmacophore map is illustrated in **Fig. 4**.

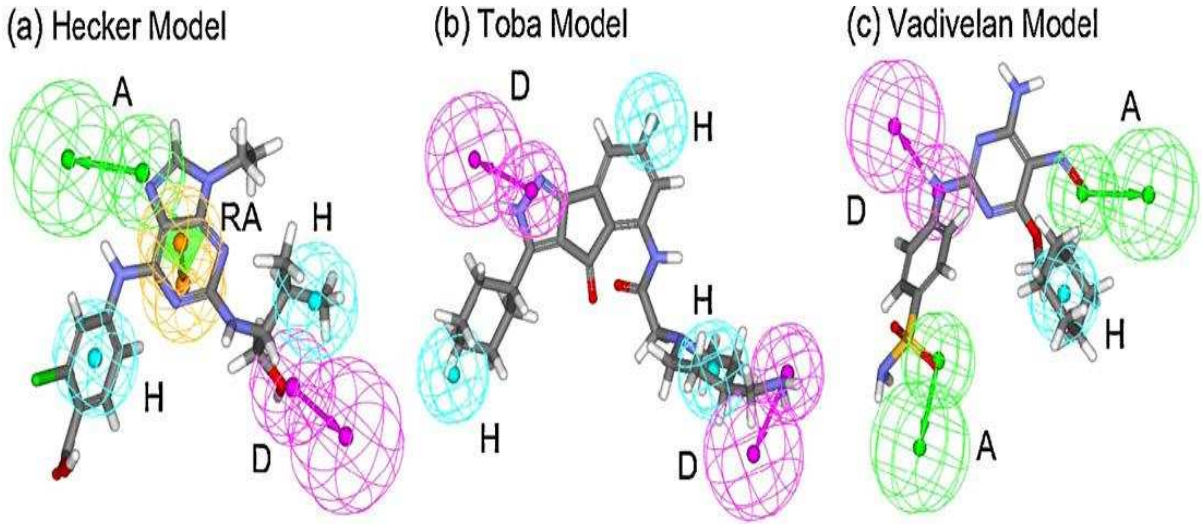


Figure 4

Flowchart of the generation of the multicomplex-based comprehensive pharmacophore map. The complex-based pharmacophore models were generated using LigandScout based on the previously aligned protein structures. All the pharmacophore features were then clustered, and the cluster centers were identified by Discovery Studio [12].

The software LigandScout 1.03, was used to generate 124 individual complex-based pharmacophore models based on the previously aligned structures. For the purposesof creating multicomplex-based comprehensive map,all the pharmacophore features identified by LigandScout were clustered according to their interaction pattern with the receptor. The software LigandScout 1.03 was used to generate 124 individual complex-based pharmacophore models based on the For the

purposes of creating multicomplex-based comprehensive map, all the pharmacophore features identified by LigandScout were clustered according to their interaction pattern with the receptor.(**figure 4**)

Hydrogen bond donor and acceptor features were clustered in term of the protein atom with which they were formed. For hydrophobic, positive and negative ionizable, and ring aromatic features, density-based clustering methods was used. The cluster centers were identified using the Accelrys Discovery Studio.

Pharmacophore model validation

Three simple 1D physicochemical properties: (1) number of hydrogen-bond acceptors, (2) number of hydrogen-bond donors, and (3) molecular weight, to ensure that for each selected inactive compound the 1D properties are similar to those of the known active compounds. The molecular structures were prepared using SciTegic Pipeline Pilot 6.1.5. Multiple conformations of each compound were generated using the FAST conformational search protocol implemented in Catalyst. All compounds were fit to the pharmacophore models using the Best Fit Compare algorithm in Catalyst.
The best-fit value of the molecule to the respective pharmacophore was calculated.

Molecules are ranked based on their fit value computed

The receiver operating characteristic (ROC) curve, in which the percentage of known CDK2 inhibitors (true positives) identified by the model is plotted against the percentage of false positives found, was generated to evaluate the performance of the pharmacophore model [13].

5.4 RESULTS AND DISCUSSION

Multicomplex-based comprehensive pharmacophore map

To develop multicomplex-based pharmacophore map 24 pharmacophoric features, including
6 hydrogen bond acceptors (A1–A6), 8 hydrogen bond donors (D1– D8), 5 hydrophobic features (H1–H5), 3 ring aromatic features (RA1–RA3), and 2 positive ionizable features (PI1 and PI2) were detected. Of these 24 detected pharmacophoric features, 7 features (A1,
A2, D1, D2, H1, H2, and H3) were found to present in the 124 complexes with more than 25% probability [12].

The most frequently occurred feature

A1 (91.94%) was mapped to the hydrogen bond acceptor of the ligand, whose partner donor is the backbone amino nitrogen of Leu83 locating the hinge region that links the two lobes of the kinase (Fig. 5a and c).
The **second** remarkable feature is **D1** (75.00%) that represents the hydrogen bond donor whose partner acceptor is the backbone carbonyl oxygen of Leu83 (Fig. 5a and c).

The **third** feature **H1** (68.55%, Fig. 5a) corresponds to the hydrophobic region formed by residues Ile10, Phe82, and Gln85 at the solvent-accessible region.

The fourth one **H2** (50.00%) represents the hydrophobic region formed by residues Val64, Phe80, and Ala144 at the small-buried region.
The **fifth** one **H3** was mapped to the hydrophobic region formed by residue Val18 in the
The **sixth** feature **D2** stands for the hydrogen bond donor interaction with the backbone carbonyl oxygen of Glu81.

The **seventh** feature **A2** represents the hydrogen bond acceptor whose partner donor is the side chain nitrogen of Lys89, locating at the solvent-accessible region that cannot be utilized by ATP (Fig. 5a and c).

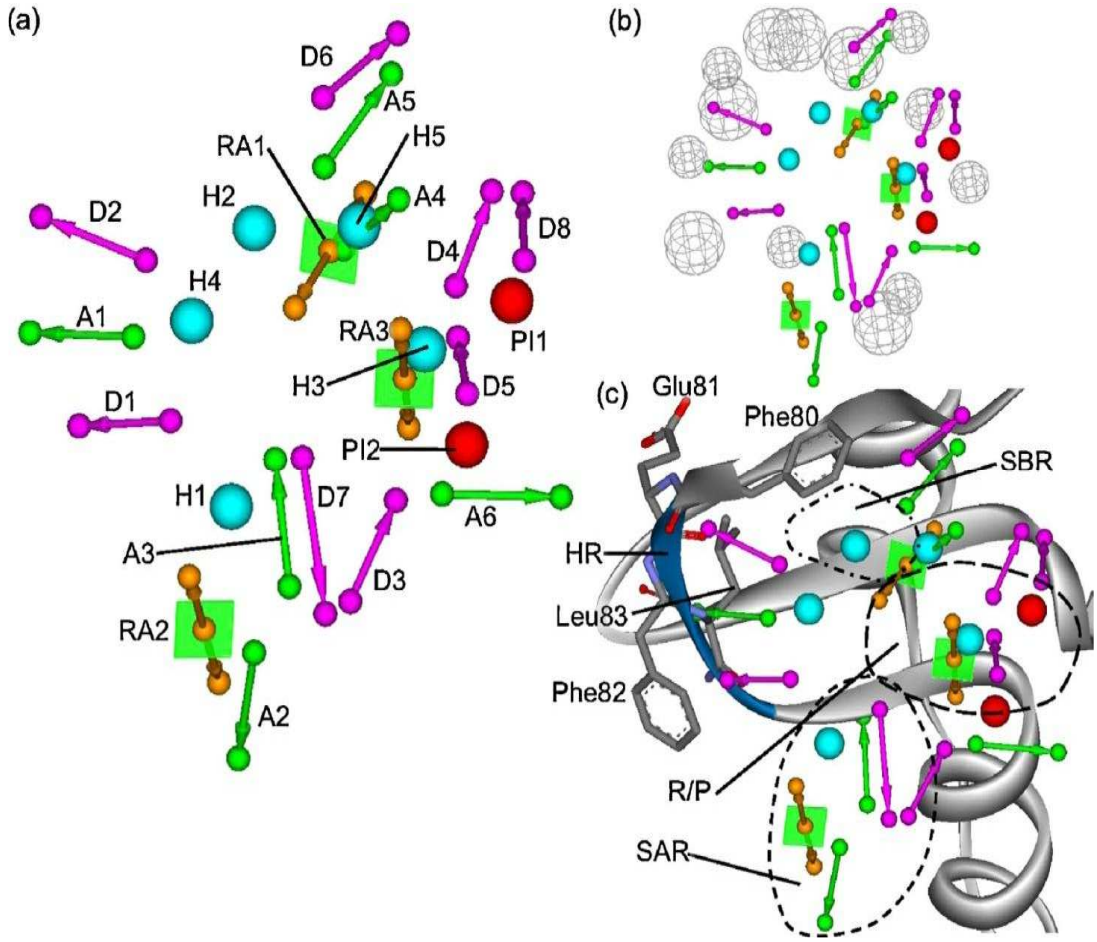


Figure.5

(a) All the chemical features of the multicomplex-based comprehensive pharmacophore map. (b) The comprehensive map with excluded volume features. For clarity some excluded volume features were not drawn. (c) View looking into the ATP-binding site of CDK2 with comprehensive map from the N-terminal domain. [12]

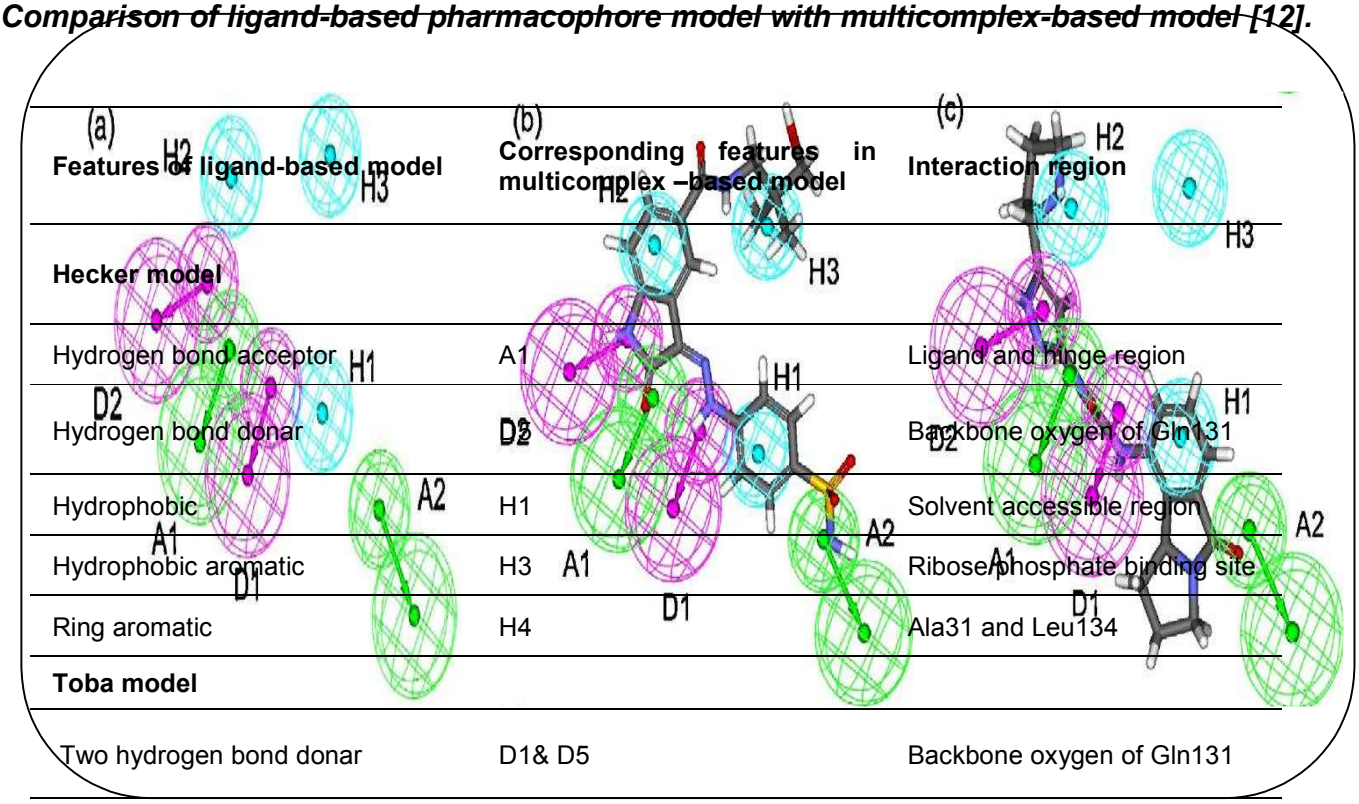
1 excluded volume features were found in the ATP-binding site
Comprehensive pharmacophore map involving excluded volume spheres has been shown in ***Fig. 5b.***

Comparisons with ligand-based pharmacophore models

The alignment of various ligand-based models with multicomplex-based model gave following inferences

Table 5

Comparison of ligand-based pharmacophore model with multicomplex-based model [12].



Features of ligand-based model	Corresponding features in multicomplex –based model	Interaction region
Hecker model		
Hydrogen bond acceptor	A1	Ligand and hinge region
Hydrogen bond donar	D2	Backbone oxygen of Gln131
Hydrophobic	H1	Solvent accessible region
Hydrophobic aromatic	H3	Ribose/phosphate binding site
Ring aromatic	H4	Ala31 and Leu134
Toba model		
Two hydrogen bond donar	D1& D5	Backbone oxygen of Gln131
Three hydrophobic	H1, H2, & H3	Ribose/phosphate binding site
Vadivelan model		
Two hydrogen bond acceptor	A3 & A4	Backbone nitrogen of Asp86
Hydrogen bond donar	D1	Amino group of Lys33
Hydrophobic	H3	

Each pharmacophore feature in the ligand based pharmacophore models; one can find a corresponding feature, indicating that the comprehensive pharmacophore map contains much more information over all the three ligand-based models. It was also noticed that some conserved features with high statistical frequency were missed in these ligand-based pharmacophore models, such as the feature D2 (statistical frequency: 38.71%) and the feature A2 (statistical frequency: 25.81%). This might limit the predictive ability of these ligand based models [12].

Most-frequent-feature pharmacophore model

the top ranked seven features (A1, A2, D1, D2, H1, H2, and H3), which were found to present in the 124 complexes with more than 25% probability [12], would be more appropriate in practice, and consequently they were selected from the comprehensive pharmacophore map and were merged to generate a most-frequent-feature pharmacophore model (**Fig. 6a**).

Figure.6

(a) The most-frequent-feature pharmacophore model. Example of (b) high active molecule Compound-079 and (c) low-active molecule Compound-092 aligned with the most-frequent-feature pharmacophore model [12].

Pharmacophore model validation

Subsequently, the most-frequent-feature pharmacophore model was validated by using it to screen against an external test dataset that includes known active and inactive compounds.

The most-frequent-feature pharmacophore model was firstly screened against external test dataset, which includes the external active dataset containing 194 known CDK2 inhibitors and the focused inactive dataset consisting of 300 non inhibitors

The receiver operating characteristic (ROC) curve [14] was used to estimate the performance of the pharmacophore model.

The best model identifies the greatest number of true positives and the least number of false positives.

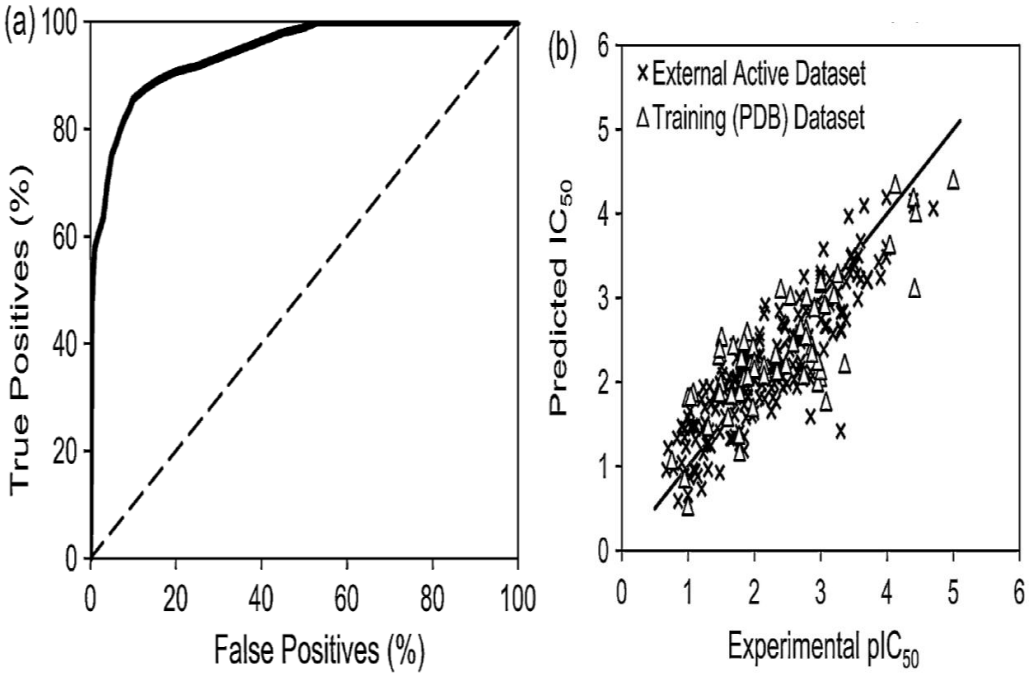


Figure.7

(a) Receiver operator characteristic (ROC) curves generated from screening the external test dataset, which includes 194 known CDK2 inhibitors and 300 inactive molecules. (b) The correlation between the experimental and predicted activities (pIC50 values) obtained from the most-frequent-feature pharmacophore model [12].

The performance of the most-frequent-feature pharmacophore model at discriminating known inhibitors versus a focused dataset of non inhibitors is shown in **Fig.** 7a. It is obvious that the most-frequent-feature pharmacophore model was very successful at differentiating between these two populations. It identifies 75.3% of the true positives and only 5.0% of the false positives.

Furthermore, in order to evaluate the efficiency of the most frequent- feature pharmacophore in virtual screening of a large database containing diverse compounds, the same validation protocol described above was applied to screen against the MDL Comprehensive Medicinal Chemistry database (CMC, version 2007.1) consisting of 8896 compounds.

This pharmacophore model displays a superior performance with 76.3% retrieval efficiency when the false-positive tolerance is only 1.0% The enrichment factor (E) calculated according to Gopalakrishnan et al. [13] was found to be 12.2, indicating that it is 12 times more probable to pick an active compound from the database than an inactive one.

The validation has shown that our model is capable of being used in virtual screening of large databases. Fig. 4b and c represents the most-frequent-feature pharmacophore model aligned with the high- and low-active molecules (Compound-079 and Compound-092) with IC50 values of 6.8 nM and 25 mM, respectively.

It is shown that the high-active Compound-079 fits all the pharmacophore features in the given relative position in space with high fit value. On the other hand, the molecule Compound- 092 with low activity does not fulfill these criteria and fits the most-frequent- feature pharmacophore

poorly and partially.

It is implied that the most-frequent-feature model provided important information on the structural and chemical properties of the CDK2 inhibitors.

6. CONCLUSIONS

It was clearly demonstrated that each one of the ligand-based pharmacophore models is just a subset of the comprehensive map. Our multicomplex-based comprehensive map has managed to involve almost all the pharmacophore features for CDK2–inhibitor interactions. It also provides us insight into how important each pharmacophore feature is.

Thus the most-frequent-feature pharmacophore model can be constructed based on the most frequent chemical features of the comprehensive pharmacophore map.The receiver operating characteristic (ROC) curve has indicated that the most-frequent- feature pharmacophore model is able to differentiate between known CDK2 inhibitors and the compounds in the focused inactive dataset

The most-frequent-feature pharmacophore model is capable of predicting the activities of a wide variety of CDK2 inhibitors in the external active dataset.

The pharmacophore model obtained can be used to retrieve potential inhibitors from a database in virtual screening and to evaluate the newly engineered compound in de novo design. The information is helpful for the study, towards more accurate pharmacophore modeling.

REFERENCES

[1] C. G. Wermuth, Pharmacophores: historical perspective and viewpoint from a medicinal chemist, Pharmacophores and pharmacophore searches, Wiley–VCH. 32 (2006) 3–13.

[2] D. O. Morgan, Cyclin-dependent kinases: engines, clocks, and microprocessors, Annu. rev.
Cell Dev. Bio. 13 (1997) 261-291.

[3] J. W. Harper, P. D. Adams, Cyclin-dependent kinases, Chem. Rev. 101 (2001) 2511-2526.

[4] M. Malumbres, M. Barbacid, Mammalian cyclin-dependent kinases, Trend Biochem. Sci. 30 (2005) 630-641.

[5] S. Wadler, Perspectives for cancer therapies with cdk2 inhibitors, Drug Resist. Update. 4 (2001)
347-367.

[6] T. M. Sielecki, J. F. Boylan, P. A. Benfield, G. L. Trainor, Cyclin-dependent kinase inhibitors:
useful targets in cell cycle regulation, J. Med. Chem. 43 (2000) 1-18.

[7] M. Johansson, J. L. Persson, Cancer therapy: targeting cell cycle regulators,Anti-Cancer Agents in Medicinal Chemistry, J. Med. Chem.. 8 (2008) 723-731.

[8] A. S. Lundberg, R. A. Weinberg, Control of the cell cycle and apoptosis, Eur. J. Cancer. 35 (1999) 1886-1894.

[9] N. M. Mascarenhas, N. Ghoshal, An efficient tool for identifying inhibitors based on 3D-QSAR and docking using feature-shape pharmacophore of biologically active conformation–A case study with CDK2/CyclinA, Eur. J. Med. Chem. 43 (2008) 2807-2818.

[10] R. Fotedar, L. Diederich, A. Fotedar, Apoptosis and the cell cycle,Progress in cell cycle research. 2 (1996) 147.

[11] S. A. Khedkar, A. K. Malde, E. C. Coutinho, S. Srivastava, Pharmacophore modeling in drug discovery and development: an overview, J. Med. Chem. 3 (2007) 187-197.

[12] J. Zou, H. Z. Xie, S. Y. Yang, J. J. Chen, J. X. Ren, Y. Q. Wei, Towards more accurate pharmacophore modeling: Multicomplex-based comprehensive pharmacophore map and most- frequent-feature pharmacophore model of CDK2,

J. Mol. Graph. Model. 27 (2008) 430-438.
[13] O. F. Guner, History and evolution of the pharmacophore concept in computer-aided drug design, Curr. Top.Med. Chem. 2 (2002) 1321-1332.
[14] S. Ekins, J. Mestres, B. Testa, In silico pharmacology for drug discovery: applications to targets and beyond, Brit. J. Pharmacol. 152 (2007) 21.